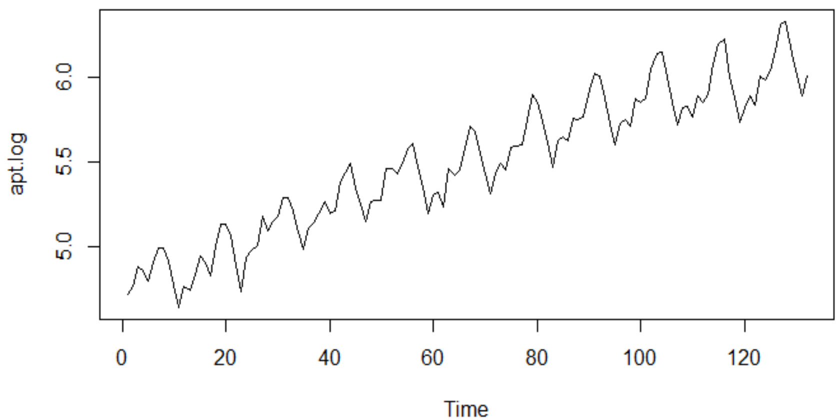
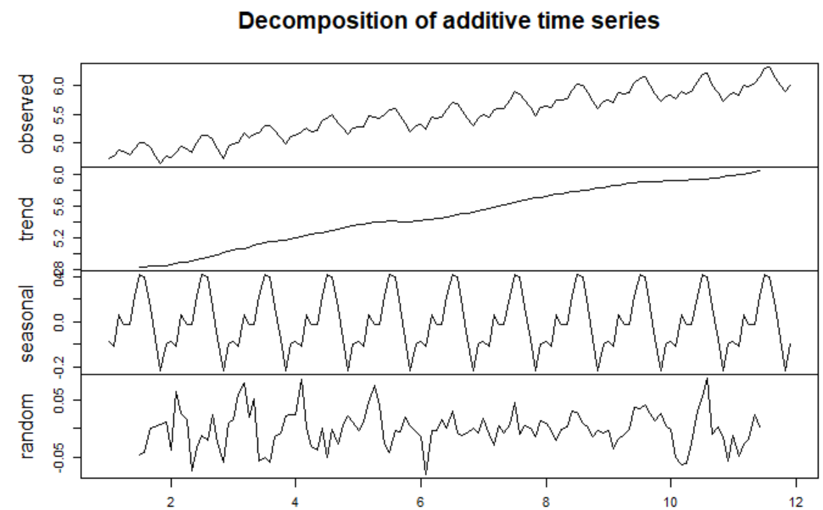
Week 4. Lecture 7: SARIMA Models.

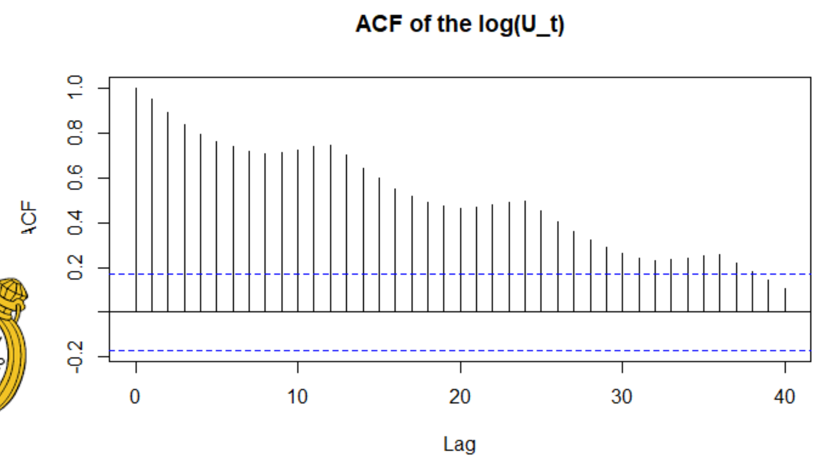
Lecture 8: Working with Data: Sample ACF/PACF.

Yule-Walker Parameter Estimation.

The Durbin-Levinson Algorithm.



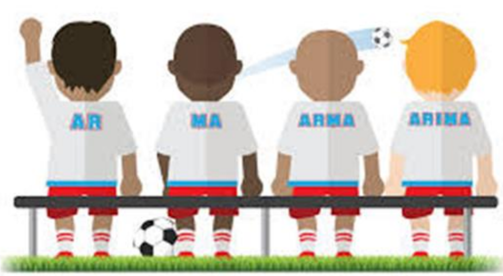


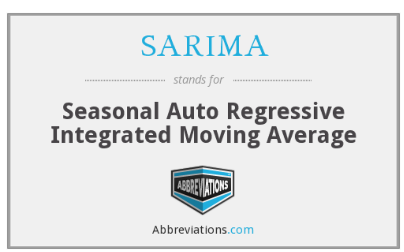


Week 4. Lecture 7: SARIMA Models.
Lecture 8: Working with Data: Sample ACF/PACF. Yule-Walker Parameter Estimation. The Durbin-Levinson Algorithm.

Outline of Lecture 7:

Outline of Lecture 8:



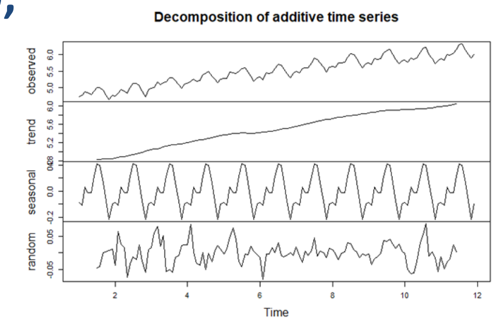


Welcome to Part I of Lecture 7:

And some of the most important concepts and notations covered in Week 3 are

Summary of the most important concepts of Lecture 6:

- PACF: the last component $\alpha(n) = \phi_{nn}$ of n-dimensional vector $\underline{\phi}_n = R_n^{-1}\,\underline{\rho}$;
- Identification of pure AR model: PACF $\alpha(n) = 0$, $n > p$; $\alpha(p) \neq 0$;
- Identification of pure MA model: ACF $\rho(n) = 0$, $n > q$; $\rho(q) \neq 0$;



- Classical Decomposition Model: $X_t = m_t + s_t + S_t$
- Differencing operators: OPERATOR ∇(nabla)

• Lag s difference: $\nabla_s X_t := X_t - X_{t-s} = (1 - B^s)\, X_t$
• Lag 1 difference: $\nabla X_t := X_t - X_{t-1} = (1 - B)\, X_t$
• dth difference at lag 1: $\nabla^d X_t = (1 - B)^d\, X_t$

- To eliminate seasonality w. period s, use lag s difference: $\nabla_s X_t$
- To eliminate polynomial trend of order d, use $\nabla^d X_t$

- Unit roots in AR part: underdifferenced
- Unit roots in MA part: overdifferenced
- Increase in variance w. differencing: overdifferenced

- Transformations: to stabilize variance; make data closer to normal

Tired of Greek letters: α, φ, ψ, ρ, ϑ, γ, σ, μ, π, and now ∇… ?

Try ancient Hebrew found in Samarian ruins; early 8th century BCE.

Using artificial intelligence, Israeli researchers concluded that inscriptions were records of shipments in Samaria. Quote:

"We achieved our results through a method comprised of image processing and newly developed statistical learning techniques."

Check your understanding

Identify the following time series model as a specific ARIMA model:

$$X_t = X_{t-1} + Z_t - 0.2\, Z_{t-1}$$

Determine which of the above statements are true:

A. AR(2)

B. MA(1)

C. ARIMA(1,1,1)

D. ARIMA(1,1,0) = ARI(1,1)

E. ARIMA (0,1,1) = IMA(1,1)

Definition: X_t follows ARIMA(p, d, q) model if

$\phi(B)\,(1-B)^d X_t = \theta(B)\, Z_t$ or $\phi^*(B)\, X_t = \theta(B)\, Z_t$

with $\phi^*(z) = \phi(z)\,(1-z)^d$ that is, *$\phi^*(z)$ has a unit root of order d;*

Roots of polynomials ϕ and θ lie outside the unit circle.

Check your answer of the next slide

Check your understanding

Identify the following time series model as a specific ARIMA model:

$$X_t = X_{t-1} + Z_t - 0.2\, Z_{t-1}$$

Determine which of the above statements are true:

A. AR(2)

B. MA(1)

C. ARIMA(1,1,1)

D. ARIMA(1,1,0) = ARI(1,1)

E. ARIMA (0,1,1) = IMA(1,1)

X_t follows ARIMA(p, d, q) model if $\phi\,(B)\,(1 - B)^d X_t = \theta\,(B)\,Z_t$ or $\phi^*\,(B)\,X_t = \theta\,(B)\,Z_t$ with $\phi^*\,(z) = \phi\,(z)\,(1 - z)^d$ that is, *$\phi^*\,(z)$ has a unit root of order d;*
Roots of polynomials ϕ and θ lie outside the unit circle.

Answers: Rewrite the model as $X_t - X_{t-1} = Z_t - 0.2\, Z_{t-1}$ or $(1\text{-}B)\, X_t = (1 - 0.2\, B)\, Z_t$.
Polynomial $\theta(z) = 1 - 0.2z$ corresponds to q=1 and has one root: $z = 5 > 1$.
Polynomial $\phi\,(z) = 1$ corresponds to p=0. This is ARIMA (0, 1, 1) as in E.
Note: $\theta(z) = 1 - 0.2z$ has one root: $z = 5 > 1$, therefore X is invertible. $\phi\,(z) = 1$ has no roots.
After differencing at lag 1 once (d=1), one gets $\nabla X_t = Z_t - 0.2\, Z_{t-1}$; it is invertible MA(1).

Check your understanding

Write equations for the following models:

A. **AR(2)**
B. **MA(1)**
C. **ARIMA(1,1,1)**
D. **ARIMA(2,1,0) = ARI(2,1)**
E. **ARIMA (0,2,1) = IMA(2,1)**

Definition : X_t **follows ARIMA(p, d, q) model if**

$$\phi(B)(1-B)^d X_t = \theta(B) Z_t \quad \text{or} \quad \phi^*(B) X_t = \theta(B) Z_t$$

with $\phi^*(z) = \phi(z)(1-z)^d$ **that is,** ***$\phi^*(z)$ has a unit root of order d;***

Roots of polynomials ϕ **and** θ **lie outside the unit circle.**

Check your answer of the next slide

Check your understanding

Write equations for the following models:

A. **AR(2):** $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t$

B. **MA(1):** $X_t = Z_t + \theta_1 Z_{t-1}$

C. **ARIMA(1,1,1):** $(1 - \phi_1 B)(1 - B) X_t = Z_t + \theta_1 Z_{t-1}$

D. **ARIMA(2,1,0) = ARI(2,1):** $(1 - \phi_1 B - \phi_2 B^2)(1 - B) X_t = Z_t$

E. **ARIMA (0,2,1) = IMA(2,1):** $(1 - B)^2 X_t = Z_t + \theta_1 Z_{t-1}$

Definition : X_t **follows ARIMA(p, d, q) model if**

$\phi(B)(1 - B)^d X_t = \theta(B) Z_t$ or $\phi^*(B) X_t = \theta(B) Z_t$

with $\phi^*(z) = \phi(z)(1 - z)^d$ **that is,** $\phi^*(z)$ ***has a unit root of order d;***

Roots of polynomials ϕ **and** θ **lie outside the unit circle.**

Check your understanding

Check your understanding:
A stationary ARMA(p, q) model is known to have zero partial autocorrelations at all lags k > 1 and nonzero autocorrelation at lag one. Determine (p, q).

Observed People :

Learned Occupations: firefighter, doctor, pirate, British soldier, AM soldier, policeman

Observed ACF/PACF patterns: ACF=0, k > q; **PACF = 0, k > p;**

ACF or PACF ≠ 0 for some h > k, for all k **ACF periodic**

Learned Models: pure MA, pure AR, mixed ARMA, (S)ARIMA

Check your answer on the next slide

Check your understanding

Check your understanding:
A stationary ARMA(p, q) model is known to have zero partial autocorrelations at all lags k > 1 and nonzero autocorrelation at lag one. Determine (p, q).

Observed
People :





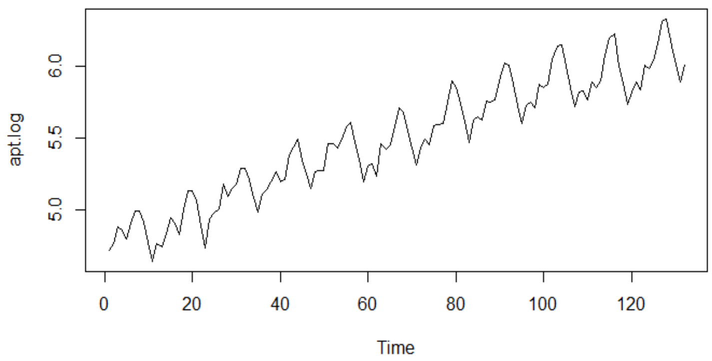

Learned
Occupations: firefighter, doctor, pirate, British soldier, AM soldier, policeman

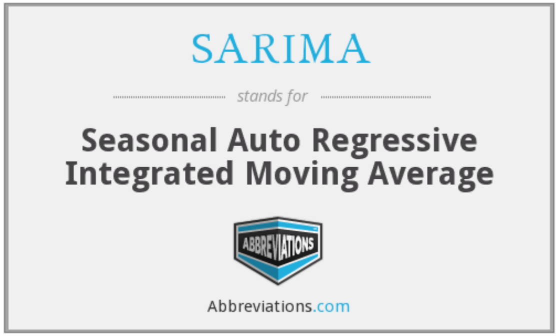
Observed ACF/PACF patterns: ACF=0, k > q; PACF = 0, k > p;
ACF or PACF ≠ 0 for some h > k, for all k ACF periodic

Learned Models: pure MA, pure AR, mixed ARMA, (S)ARIMA

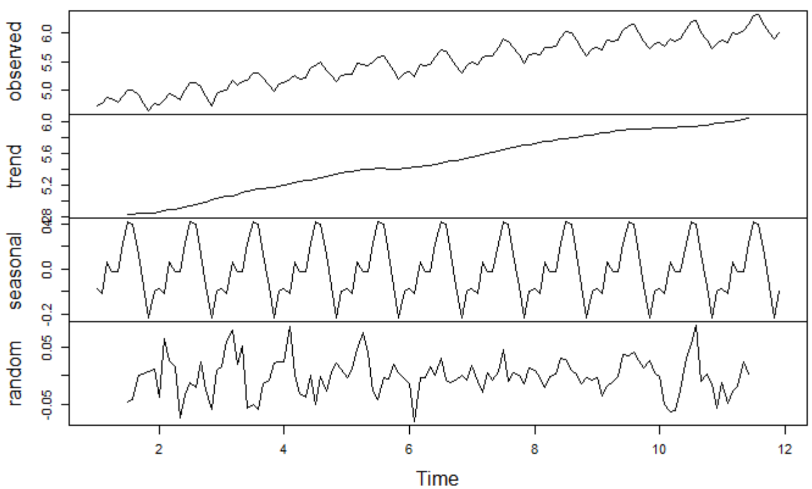
Answer: AR(1) model is the only ARMA model with the given property: PACF $\alpha(k) = 0$ for $k > 1$, and $\rho(1) \neq 0$. Therefore, it is AR(1).

Welcome to PART II of Lecture 7: SARIMA Models.

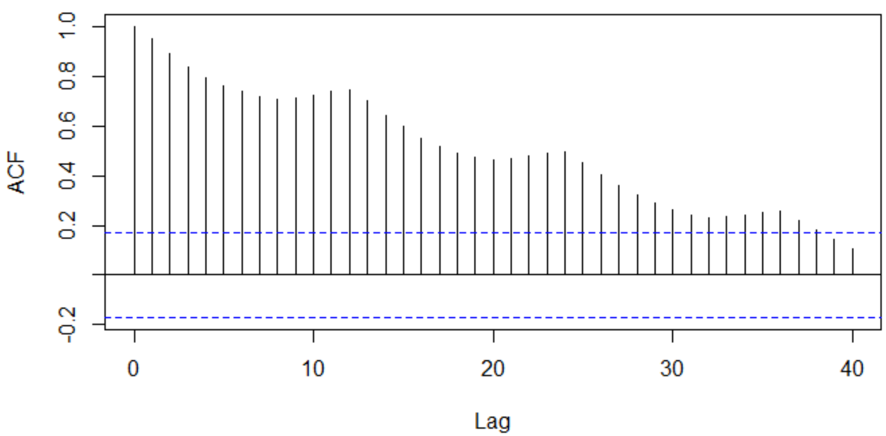




Decomposition of additive time series



ACF of the log(U_t)

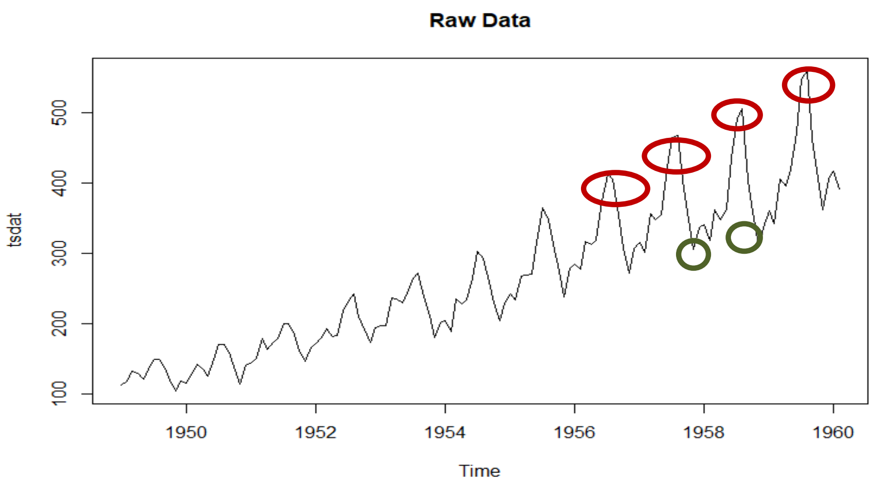


Outline of Part II of Lecture 7:

9. Seasonal ARIMA, s=12, Intuition.

January 1980 X_1	February 1980 X_2	March 1980 X_3	...	December 1980 X_{12}
January 1981 X_{13}	February 1981 X_{14}	March 1981 X_{15}	...	December 1981 X_{24}
...	...	...	...	...
January 2016 X_{433}	February 2016 X_{434}	March 2016 X_{435}	...	December 2016 X_{444}
$\downarrow$ $X_1, X_{13}, X_{25}, \ldots$	$\downarrow$ $X_2, X_{14}, X_{26}, \ldots$	$\downarrow$ $X_3, X_{15}, X_{27}, \ldots$	$\downarrow$ ...	$\downarrow$ ...

We thus have a total of s=12 series (only January or only February, etc), each has r=37 entries.



- **View time series X_t as s series:**
- **X_j, X_{j+s}, ..., $X_{j+(r-1)s}$, j=1, 2, ..., 12 = s.**

- **for January, j = 1: X_1, X_{13}, X_{25}, ...**
- **for February, j = 2: X_2, X_{14}, X_{26}, ...**
- **for March, j = 3: X_3, X_{15}, X_{27}, ...**
- **...**
- **for December, j = 12: X_{12}, X_{24}, X_{36}, ...**

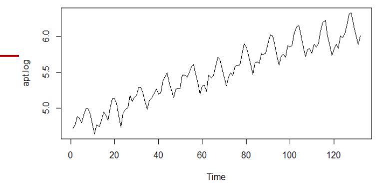
Consider: **SARIMA (p,d,q) x (P, D, Q)$_s$**

9. Seasonal ARIMA, s=12, Assumptions.

January 1980 X_1	February 1980 X_2	March 1980 X_3	...	December 1980 X_{12}
January 1981 X_{13}	February 1981 X_{14}	March 1981 X_{15}	...	December 1981 X_{24}
...	...	...	...	...
January 2016 X_{433}	February 2016 X_{434}	March 2016 X_{435}	...	December 2016 X_{444}
$\downarrow$ $X_1, X_{13}, X_{25}, \ldots$	$\downarrow$ $X_2, X_{14}, X_{26}, \ldots$	$\downarrow$ $X_3, X_{15}, X_{27}, \ldots$	$\downarrow$ ...	$\downarrow$ ...

We thus have a total of s=12 series (only January or only February, etc), each has r=37 entries.

SARIMA (p,d,q) x (P, D, Q)$_s$



Main points:

- **The model is for a data set with stable variance. If necessary, transform data.**

A. Difference D times at lag s to remove seasonality.

B. Difference d times at lag 1 to remove trend.

- **At this point the process should be stationary.**

C. Consider a dataset of annual entries for the same month, i.e., July every year. It is a stationary time series. Model it by ARMA(P,Q).

- **Assumption: annual data for each month, is modelled by the same ARMA(P,Q) .**

That is, a data set of annual entries for month of January will be modelled by the same ARMA(P,Q) as the dataset of annual entries for the month of July.

D. Consider a dataset of s = 12 entries for one year, e.g., Jan-Dec of 1980. This is a stationary dataset, model it by ARMA(p,q).

- **Assumption: for each year, 1980, 1981, etc., within each year, data follows ARMA(p,q)**

9. Seasonal ARIMA, s=12. Between-year and within-year Models.

January 1980 X_1	February 1980 X_2	March 1980 X_3	...	December 1980 X_{12}
January 1981 X_{13}	February 1981 X_{14}	March 1981 X_{15}	...	December 1981 X_{24}
...	...	...	...	...
January 2016 X_{433}	February 2016 X_{434}	March 2016 X_{435}	...	December 2016 X_{444}
$\downarrow$ $X_1, X_{13}, X_{25}, \ldots$	$\downarrow$ $X_2, X_{14}, X_{26}, \ldots$	$\downarrow$ $X_3, X_{15}, X_{27}, \ldots$	$\downarrow$ ...	$\downarrow$ ...

We thus have a total of s=12 series (only January or only February, etc), each has r=37 entries.

- **View time series X_t as s series: X_j, X_{j+s}, ..., $X_{j+(r-1)s}$, j=1, 2, ..., s**

  -- for January: j = 1; for February: j = 2; ..., ; for December : j = 12 = s. --

- **Model Assumption 1: for each j, the series is generated by the same ARMA(P,Q):**

$$(1 - \Phi_1 B^s - \ldots - \Phi_P B^{sP})X_t = (1 + \Theta_1 B^s + \ldots + \Theta_Q B^{sQ})U_t,$$

(because for $t = j + s\tau$, $B^{sP}X_t = X_{(j+s\tau)-sP} = X_{j+s(\tau-P)}$)

Between-Year Model Summary: $\Phi(B^s)X_t = \Theta(B^s)U_t,$

- **Model Assumption 2: dependence within each year for follows the same ARMA(p,q) :**

$$\phi(B)U_t = \theta(B)Z_t,\ Z_t \sim WN(0, \sigma_Z^2).$$

9. Seasonal ARIMA: Summary.

- **View time series X_t as s series: X_j, X_{j+s}, …, $X_{j+(r-1)s}$, j = 1, 2, …,**

  -- for January: j = 1; for February: j = 2; …, ; for December : j= 12 = s. --

- **Model Assumption 1: for each j, the series is generated by the same ARMA(P,Q)**

$$(1-\Phi_1 B^s - \ldots - \Phi_P B^{sP})X_t = (1+\Theta_1 B^s + \ldots + \Theta_Q B^{sQ})U_t,$$

(because for $t = j + s\tau$, $B^{sP}X_t = X_{(j+s\tau)-sP} = X_{j+s(\tau-P)}$)

Between-Year Model Summary: $\Phi(B^s)X_t = \Theta(B^s)U_t,$

- **Model Assumption 2: dependence within each year follows the same ARMA(p,q)**

$$\boxed{\phi(B)U_t = \theta(B)Z_t,\ Z_t \sim WN(0,\sigma_Z^2).}$$

- **Summary:**

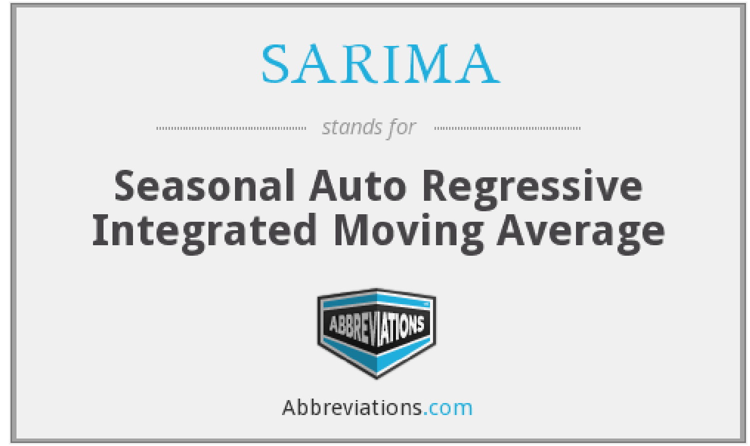
SARIMA (p,d,q) x (P, D, Q)$_s$

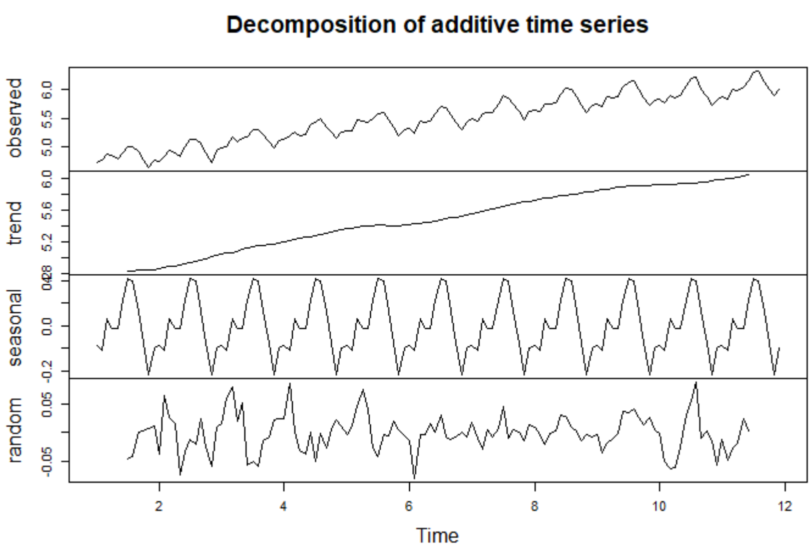
$$\boxed{\phi(B)\Phi(B^s)Y_t = \theta(B)\Theta(B^s)Z_t,\ \ Z_t \sim WN(0,\sigma_Z^2),\ for\, Y_t := (1-B)^d(1-B^s)^D X_t.}$$

(Usual assumptions on polynomials)

Continue to Examples of SARIMA models in Part III of Lecture 7

Welcome to PART III of Lecture 7: SARIMA Models.





Outline of Part III of Lecture 7:

Examples: SMA(1)$_s$ & SAR(1)$_s$: pp. 18 - 22

Example: mixed p,q, P, Q: pp. 23 - 25

Summary: P/ACF behavior: p. 26 - 27

How to identify P, Q, p, q: p. 27

9.1 - 9.2 Examples SARIMA $(0,0,0) \times (P,0,Q)_{12}$

SARIMA $(p,d,q) \times (P, D, Q)_s$

$$\phi(B)\Phi(B^s)Y_t = \theta(B)\Theta(B^s)Z_t, \;\; Z_t \sim WN(0, \sigma_Z^2), \; for\, Y_t := (1-B)^d(1-B^s)^D X_t.$$

9.1 Example: SARIMA $(0,0,0) \times (1,0,0)_{12}$ also denoted SAR $(1)_{12}$

$s=12$; $d=D=0$; $p = q = 0$, $P = 1$, $Q = 0$. Thus, $\Phi(z) = 1 - \Phi_1 z$ and the model is

$(1 - \Phi_1 B^{12}) X_t = Z_t$ or $X_t - \Phi_1 X_{t-12} = Z_t$

9.2 Example: SARIMA $(0,0,0) \times (0,0,1)_{12}$ also denoted SMA $(1)_{12}$

$s=12$; $d=D=0$; $p=q=0$, $P = 0$, $Q = 1$. Thus, $\Theta(z) = 1 + \Theta_1 z$ and the model is

$X_t = (1 + \Theta_1 B^{12}) Z_t$ or $X_t = Z_t + \Theta_1 Z_{t-12}$.

$\gamma_Z(k) = E(Z_t Z_{t+k}) = 0, \; k \neq 0$
$\gamma_Z(0) = E(Z_t^2) = \sigma_Z^2$

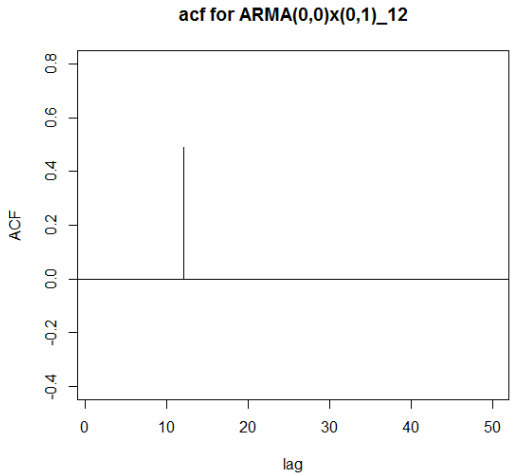
Calculate ACVF:

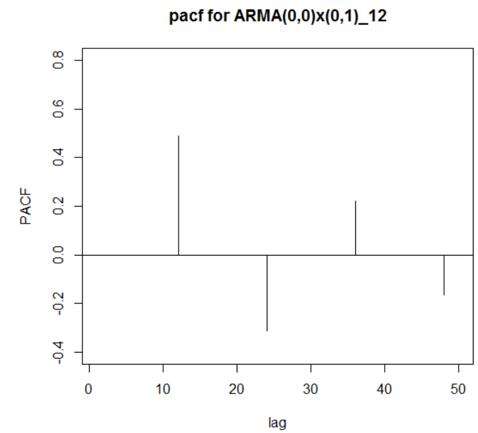
$\gamma_X(k) = E\{(Z_t + \Theta_1 Z_{t-12})(Z_{t+k} + \Theta_1 Z_{t+k-12})\} = \gamma_Z(k) + \Theta_1\{\gamma_Z(k-12) + \gamma_Z(k+12)\} + \Theta_1{}^2\gamma_Z(k)$

SMA$(1)_{12}$: $\gamma_X(k) \neq 0$ iff $k=0$ or $k = \pm 12$.

ACF/PACF for SMA $(1)_{12}$

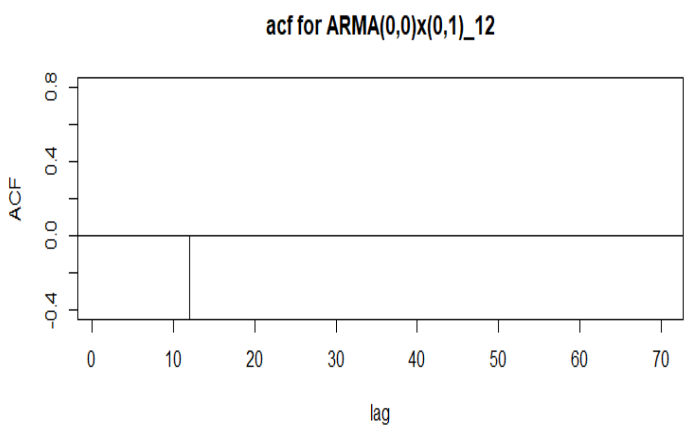
Recall from Example 6.1, Week 3: for MA(1), k >0
$\alpha(k) = \phi_{kk} = (-1)^{(k+1)}\theta_1{}^k/(1+\theta_1{}^2+\ldots+\theta_1{}^{2k})$.

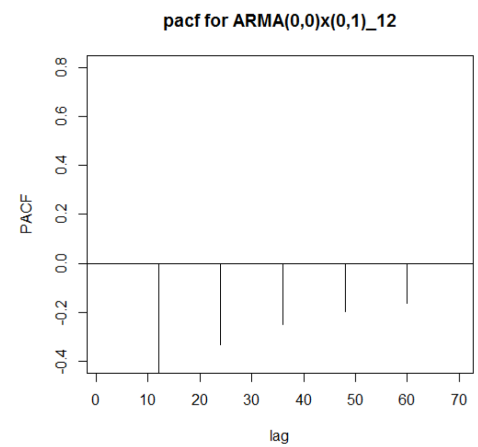




$\Theta > 0$

ACF/PACF for SMA(1)

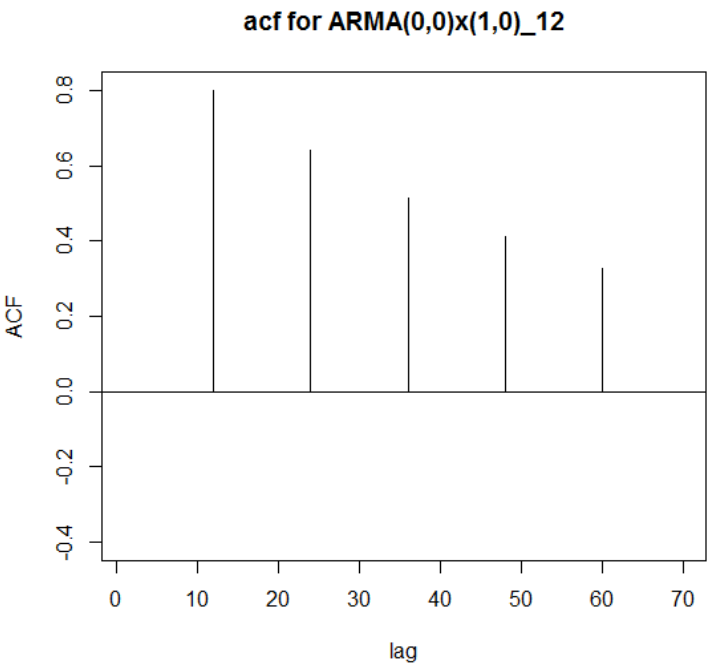


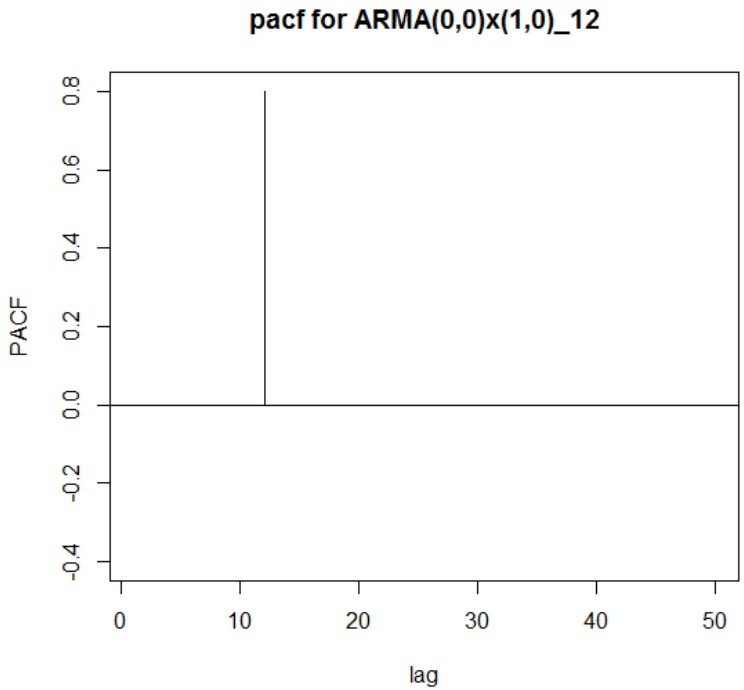


$\Theta < 0$

SMA(1) $_{12}$: ACF $\rho_X(k) \neq 0$ $only\ for$ k=0 or k = $\pm$ 12.

Theoretical acf/pacf for SAR(1)$_{12}$: $X_t - 0.8X_{t-12} = Z_t$.





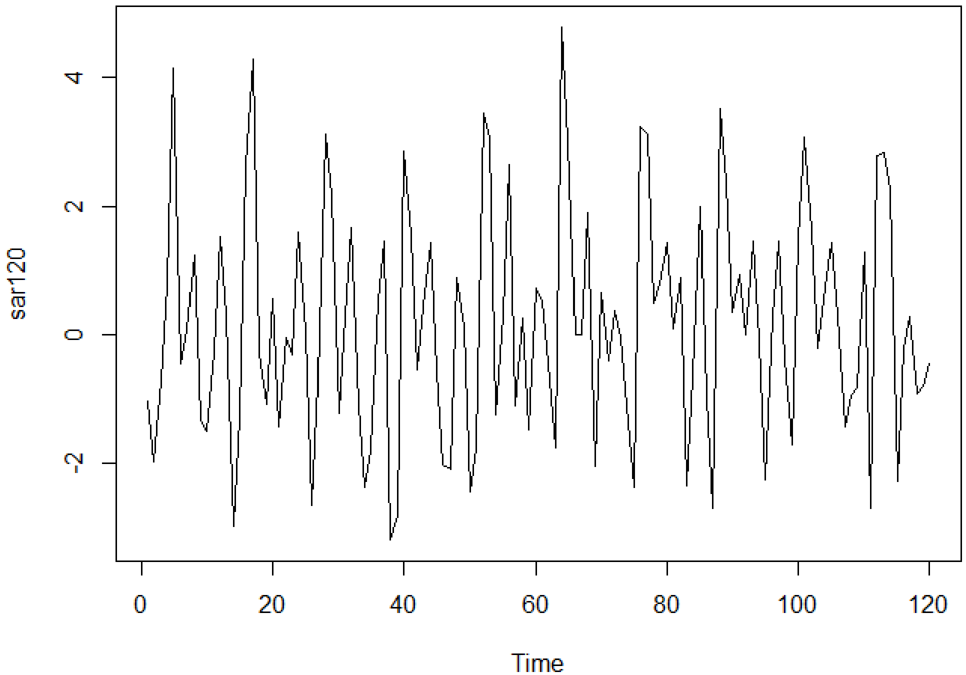
Annual AR: $\rho(12k) = \Phi^k$, k=1,2,…;   Identification via PACF!

Theoretical ACF looks like exponentially decaying spikes at lags 12, 24, 36, etc.

R code to plot theoretical acf and pacf for pure SAR: X_t-0.8 X_{t-12}=Z_t (seasonal with s=12, Φ_1=0.8):

```
> phi=c(rep(0,11), 0.8)                                  #pure seasonal AR(1) with phi=0.8
> ACF=ARMAacf(ar=phi,ma = 0, 70) [-1]                    #[-1]to remove lag 0
> PACF=ARMAacf(ar=phi,ma = 0, 50, pacf=TRUE)
> plot(ACF, type="h", xlab="lag", ylim=c(-.4, .8), main="acf for ARMA(0,0)x(1,0)_12"); abline(h=0)
> plot(PACF, type="h", xlab="lag", ylim=c(-.4, .8), main="pacf for ARMA(0,0)x(1,0)_12"); abline(h=0)
```

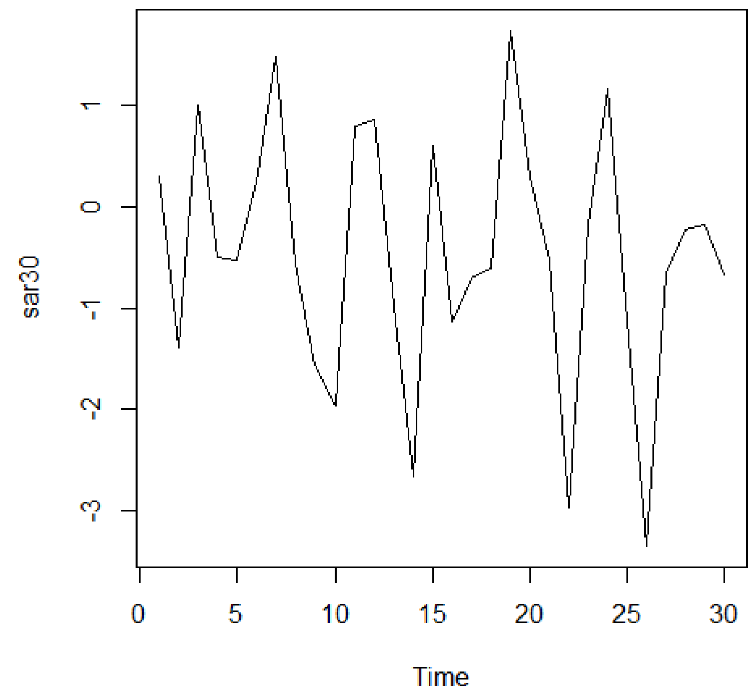
Simulated 120 values of $SAR(1)_{12}$: $X_t - 0.8X_{t-12} = Z_t$ (P=1, Q=0)



SAR(1)

10 periods: n=12 x 10

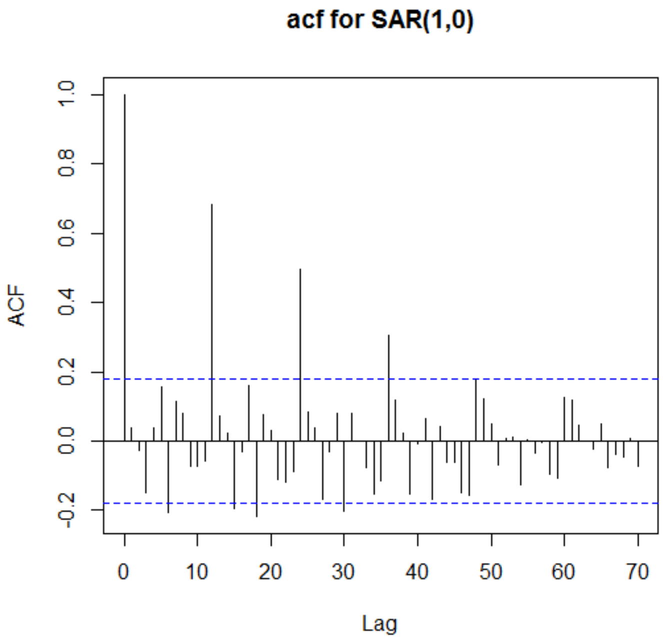
3 periods: n=12 x 3

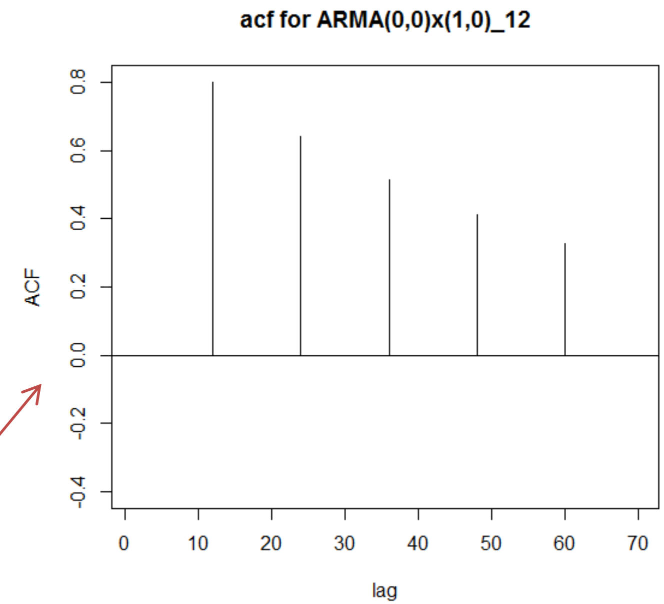


R code to simulate data:

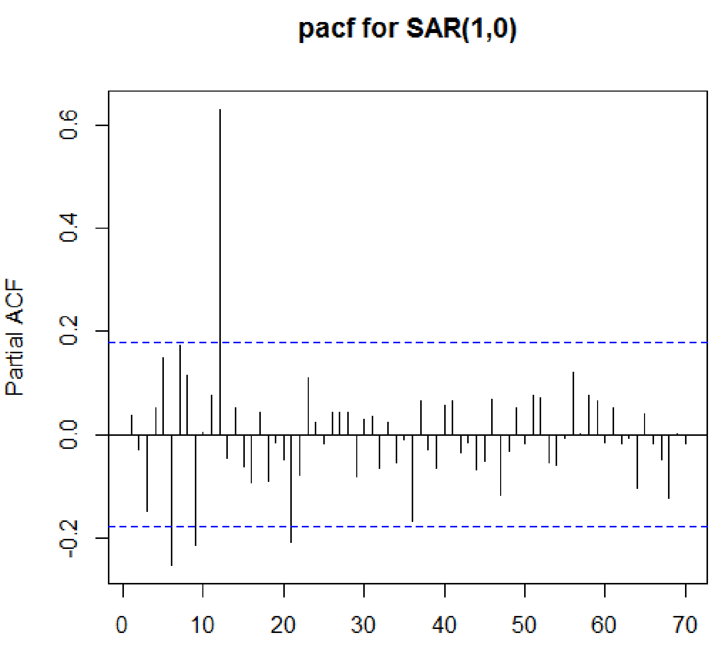
```
> set.seed(90210)
> phi=c(rep(0,11), 0.8)
> sar120 <- arima.sim(list(ar=phi), n = 120, sd = 1)
```

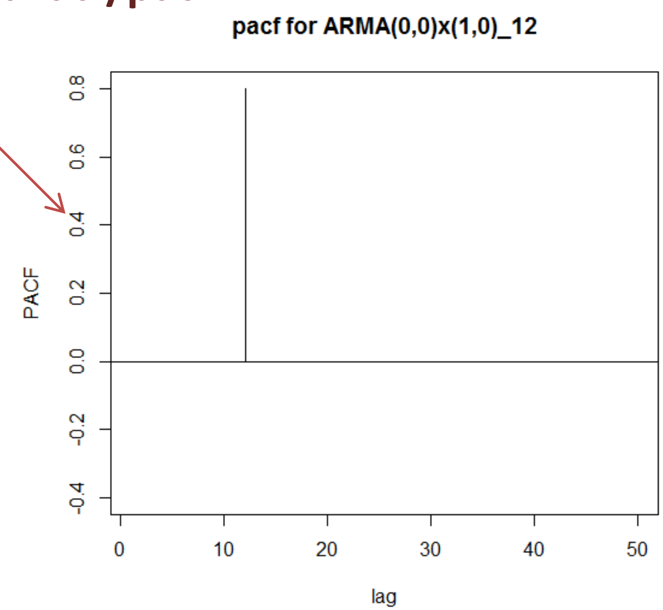
Sample acf/pacf for the simulated model: $X_t - 0.8X_{t-12} = Z_t$





theoretical acf/pacf





9.3 Example SARIMA $(0,0,1) \times (0,0,1)_{12}$

SARIMA $(p,d,q) \times (P, D, Q)_s$

$$\phi(B)\Phi(B^s)Y_t = \theta(B)\Theta(B^s)Z_t,\ \ Z_t \sim WN(0,\sigma_Z^2),\ for\, Y_t := (1-B)^d(1-B^s)^D X_t.$$

9.3 Example: SARIMA $(0,0,1) \times (0,0,1)_{12}$ (s=12; d=D=0; p=P=0, q=Q=1)

Model: $X_t = (1 + \theta_1 B)(1 + \Theta_1 B^{12}) Z_t = Z_t + \theta_1 Z_{t-1} + \Theta_1 Z_{t-12} + \theta_1 \Theta_1 Z_{t-13}$.

MA(13) model $\rightarrow \rho_X(k) = 0,\ k > 13$.

Note the last coefficient: Very special MA(13)

ACF is non-zero at lags:

$E(Z_t Z_s) = 0,\ t \neq s.$

Lag 1: $\gamma_X(1) = E(X_t X_{t-1}) = E\{(\ldots + \theta_1 Z_{t-1} + \ldots)(Z_{t-1} + \ldots)\} \neq 0$

Lag 12: $\gamma_X(12) = E(X_t X_{t-12}) = E\{(\ldots + \Theta_1 Z_{t-12} + \ldots)(Z_{t-12} + \ldots)\} \neq 0$

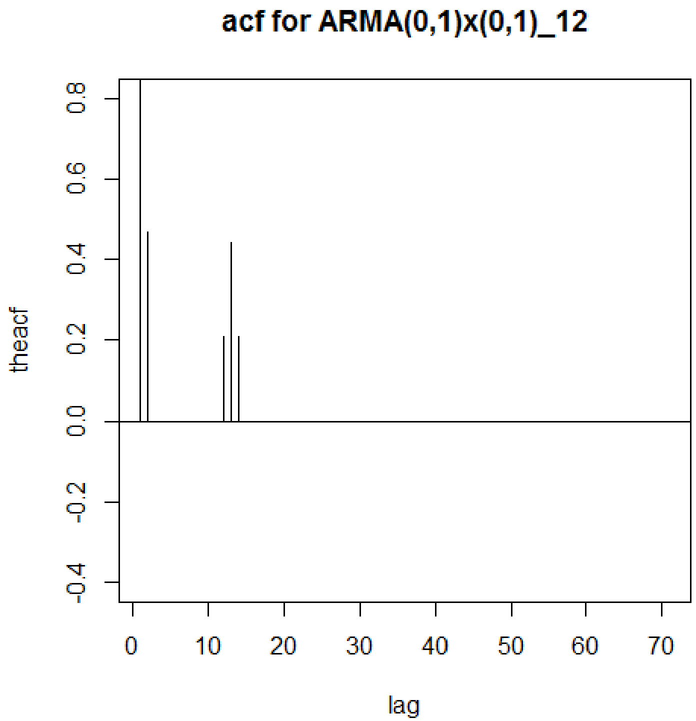
Lag 13: $\gamma_X(13) = E(X_t X_{t-13}) = E\{(\ldots + \theta_1 \Theta_1 Z_{t-13})(Z_{t-13} + \ldots)\} \neq 0$

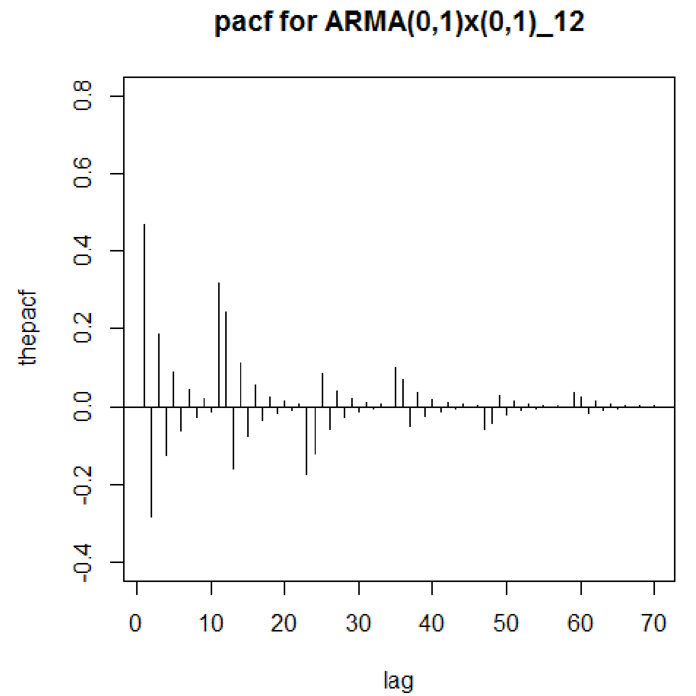
ALSO

Lag 11: $\gamma_X(11) = E(X_t X_{t-11}) = E\{(\ldots + \Theta_1 Z_{t-12} + \ldots)(Z_{t-11} + \theta_1 Z_{t-12} + \ldots)\} \neq 0$

$SMA(1)_{12}$: $ACF \neq 0$ for lags 0 and 12.
SARIMA, q=Q=1: ACF $\neq 0$ for lags 0, q=1, s-1 = 11, s = 12, s+1 = 13.

ACF/PACF for SARIMA $(0,0,1)\times(0,0,1)_{12}$



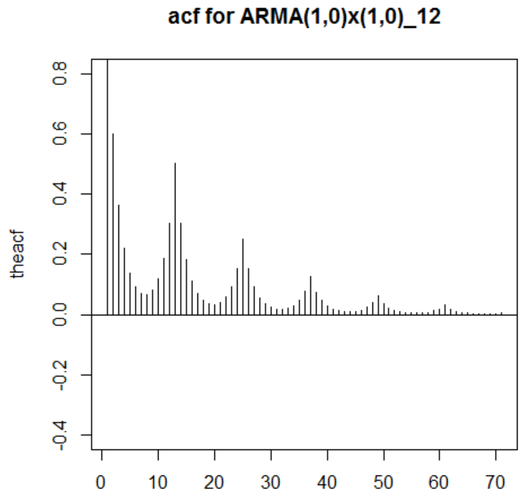


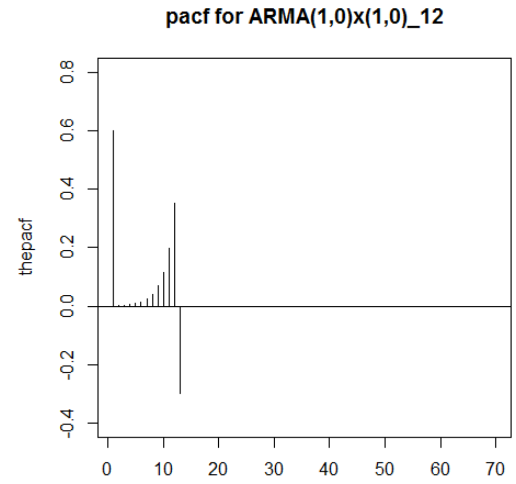
ACF/PACF for SARIMA$(0,0,1)\times(0,0,1)_{12}$: SMA(1) in between-year model; MA(1) in between-month model

SARIMA(0,0,1)x(0,0,1): $Y_t=(1+.7B)(1+.6B^{12})Z_t = Z_t + .7Z_{t-1} +.6\, Z_{t-12} + .42\, Z_{t-13}$

```
> theacf=ARMAacf (ma =c(.7,0,0,0,0,0,0,0,0,0,0,.6,.42),lag.max=70)
> plot(theacf, type="h", xlab="lag", ylim=c(-.4, .8), main="acf for ARMA(0,1)x(0,1)_12"); abline(h=0)
> thepacf=ARMAacf (ma = c(.7,0,0,0,0,0,0,0,0,0,0,.6,.42),lag.max=70, pacf=T)
> plot(thepacf, type="h", xlab="lag", ylim=c(-.4, .8), main="pacf for ARMA(0,1)x(0,1)_12"); abline(h=0)
```

9.4 More ACF/PACF for SARIMA $(1,0,0)x(1,0,0)_{12}$





Example 9.4: Let $Q = q = 0, P = p = 1, s = 12$. SARIMA$(1,0,0) \times (1,0,0)_{12}$ model for Y_t:

$$(1 - \phi_1 B)(1 - \Phi_1 B^{12})Y_t = Z_t, \text{ or}$$

$$(1 - \phi_1 B - \Phi_1 B^{12} + \phi_1\Phi_1 B^{13})Y_t = Z_t \text{ or } Y_t - \phi_1 Y_{t-1} - \Phi Y_{t-12} + \phi_1\Phi_1 Y_{t-13} = Z_t$$

For example, if $\phi = .6$ and $\Phi = .5$ we have: $Y_t - .6Y_{t-1} - .5Y_{t-12} + .3Y_{t-13} = Z_t$.

This is an AR(13) model so that PACF $\alpha(h) = 0$ for $h > 13$.
PACF has distinct spikes at lags 1, 12, 13 with a bit of action coming before lag 12.

R commands to generate these graphs:

```
> theacf=ARMAacf (ar = c(.6,0,0,0,0,0,0,0,0,0,0,.5,-.30),lag.max=70)
> plot(theacf, type="h", xlab="lag", ylim=c(-.4, .8), main="acf for ARMA(1,0)x(1,0)_12"); abline(h=0)
> thepacf=ARMAacf (ar = c(.6,0,0,0,0,0,0,0,0,0,0,.5,-.30),lag.max=70,pacf=T)
> plot(thepacf, type="h", xlab="lag", ylim=c(-.4, .8), main="pacf for ARMA(1,0)x(1,0)_12"); abline(h=0)
```

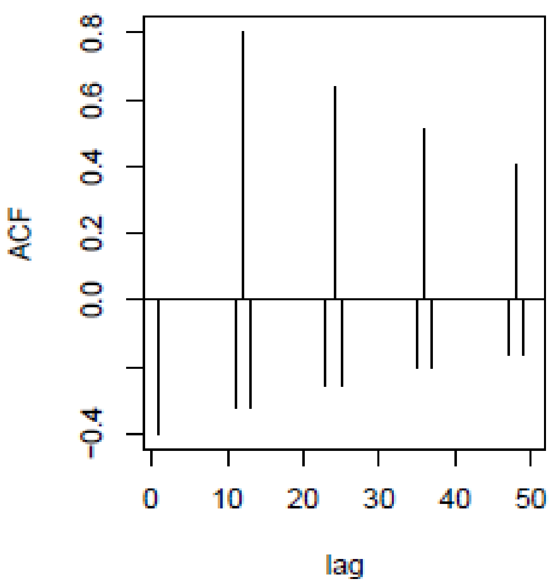
Summary: Behavior of the ACF and PACF for Pure SARMA Models

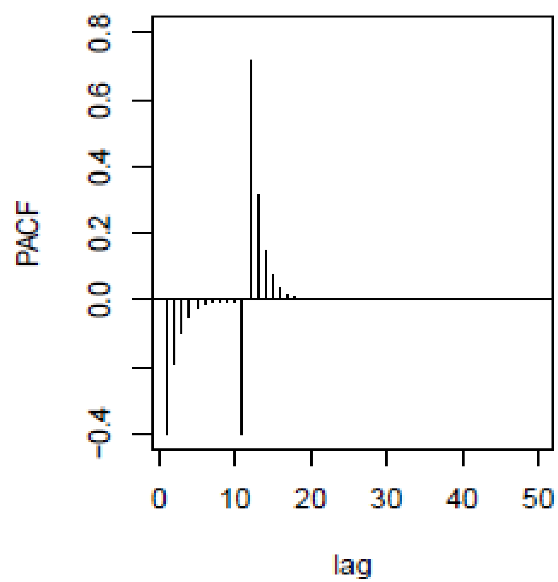
	$AR(P)_s$	$MA(Q)_s$	$ARMA(P,Q)_s$
ACF*	Tails off at lags ks, $k = 1, 2, \ldots,$	Cuts off after lag Qs	Tails off at lags ks
PACF*	Cuts off after lag Ps	Tails off at lags ks $k = 1, 2, \ldots,$	Tails off at lags ks

*The values at nonseasonal lags $h \neq ks$, for $k = 1, 2, \ldots,$ are zero.

ACF and PACF of the mixed seasonal ARMA model $X_t - 0.8 X_{t-12} = Z_t - 0.5 Z_{t-1}$

Here P=1, Q=0, p=0, q=1, s=12: $(1 - 0.8B^{12}) X_t = (1 - 0.5B) Z_t$





Summary: Seasonal ARIMA, s=12

- **View time series X_t as s series: X_j, X_{j+s}, …, $X_{j+(r-1)s}$, j=1, 2, …, s**
  -- for January: j = 1; for February: j=2; …, ; for December : j=12=s.—

- **Model Assumption 1: for each j, the series is generated by the same ARMA(P,Q)**
- **Model Assumption 2: dependence within each year for follows the same ARMA(p,q):**

SARIMA (p,d,q) x (P, D, Q)$_s$

$$\phi(B)\Phi(B^s)Y_t = \theta(B)\Theta(B^s)Z_t,\ \ Z_t \sim WN(0,\sigma_Z^2),\ for\, Y_t := (1-B)^d(1-B^s)^D X_t.$$

<u>Procedure to identify SARIMA:</u>

1. Find d, D to make $Y_t = (1-B)^d(1-B^s)^D X_t$ stationary. In practice usually use $d = 1, 2$ and $D = 1$.

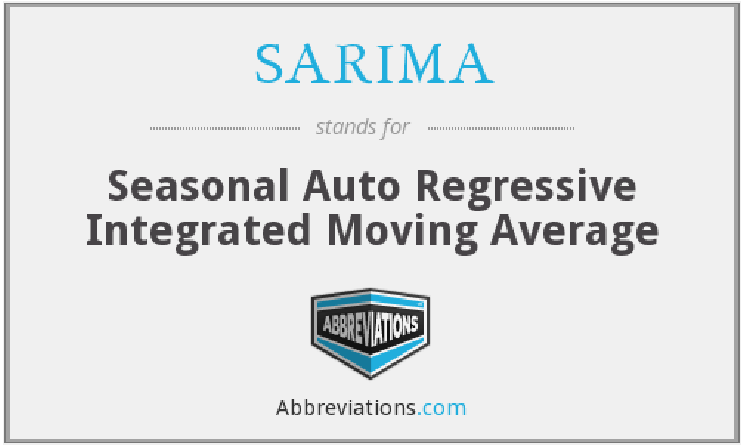
2. Find P and Q: look at $\hat{\rho}(ks)$, $k = 1, 2, \ldots$, i.e. look at ACF and PACF at lags which are multiples of s. Identify ARMA(P, Q).

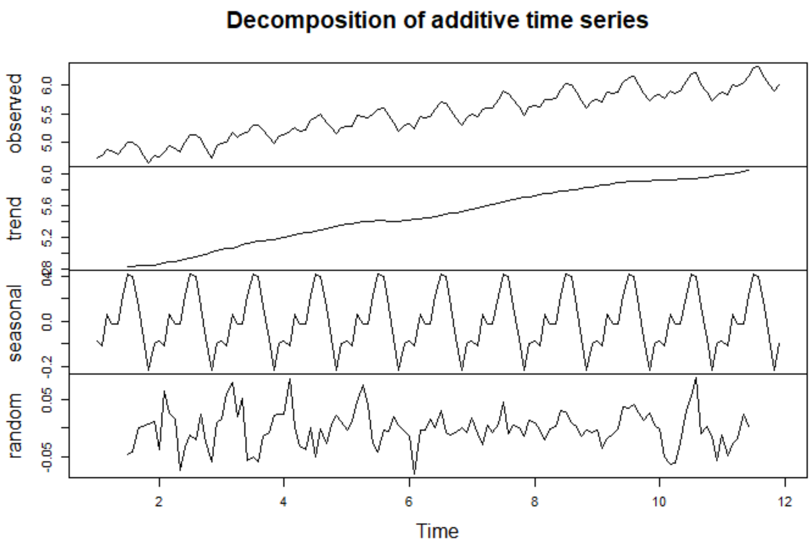
3. Find p, q: $\hat{\rho}(1), \ldots, \hat{\rho}(s-1)$ should look as ACF of ARMA (p, q).

Note: Y_t constitutes ARMA$(p + sP, q + sQ)$ process in which some of the coefficients are zeros and the rest of the coefficients are functions of $\underline{\beta}' = (\underline{\phi}', \underline{\Phi}', \underline{\theta}', \underline{\Theta}')$.

4. Use ML Estimation for $(\underline{\beta}, \sigma_Z^2)$ and use AICC and diagnostic checking to identify the best model.

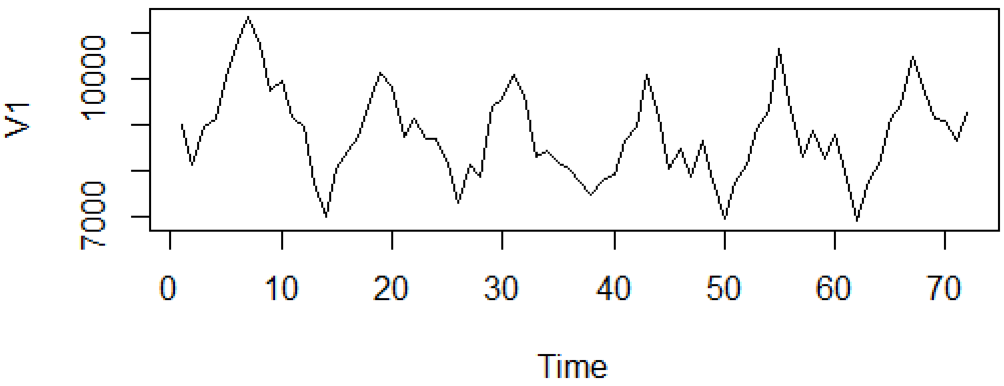
Welcome to PART IV of Lecture 7: SARIMA Models.



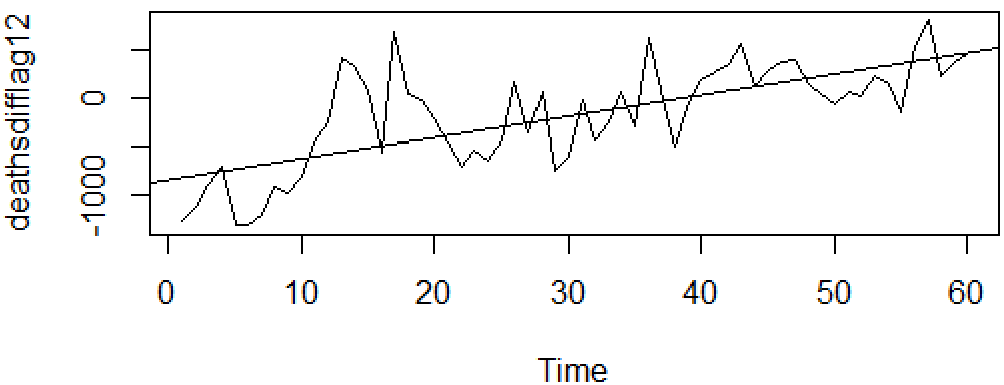


Outline of Part IV of Lecture 7:

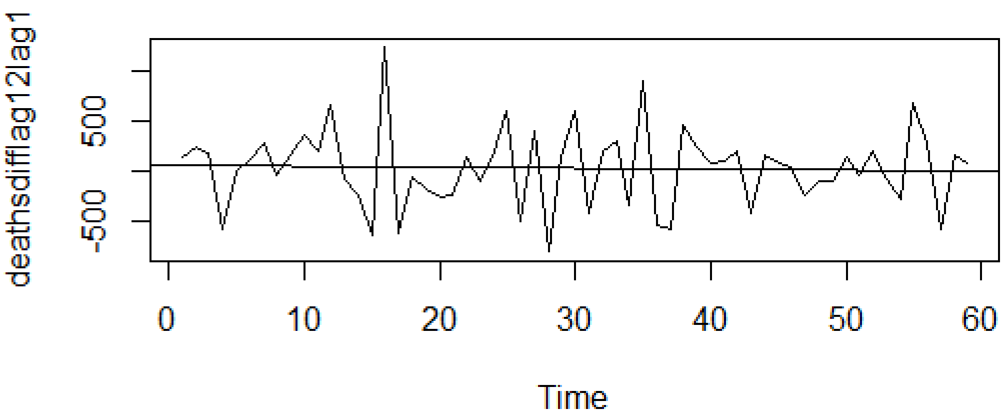
Accidental Deaths Data, differenced at lag 12 and then at lag 1



Original data: the monthly accidental deaths, 1973-1978
Sample variance: 918,411.7

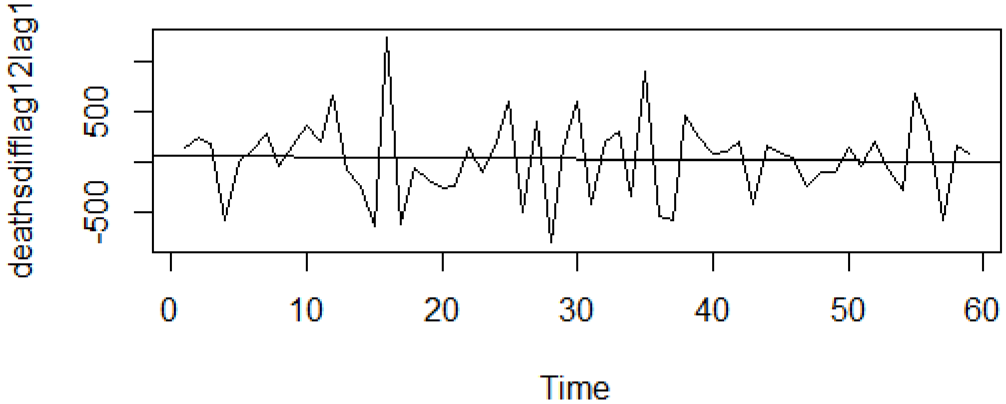


Difference at lag 12 to remove seasonality
Sample variance: 288,714.5

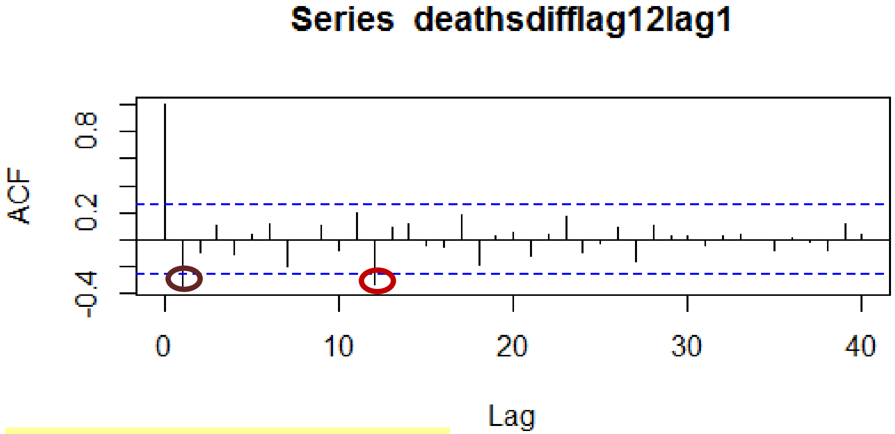


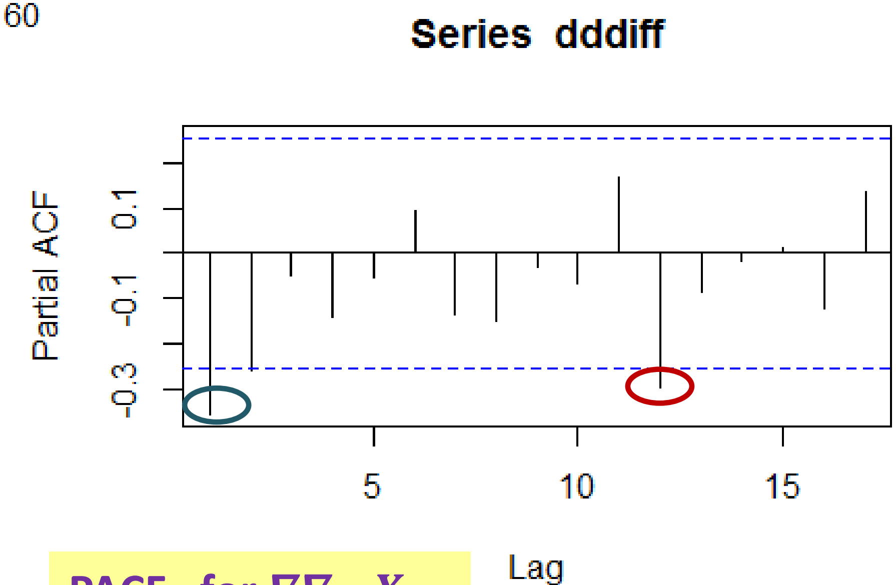
Differenced again at lag 1 to remove trend: $\nabla\nabla_{12} X_t$
Sample variance: 155,301.9

ACF /PACF for Differenced Accidental Deaths Data



Data Differenced at lag 12 and again at lag 1 to remove trend: $\nabla\nabla_{12}X_t$





ACF for $\nabla\nabla_{12}\ X_t$

PACF for $\nabla\nabla_{12}X_t$

Observations: **In SARIMA D = 1, d = 1**

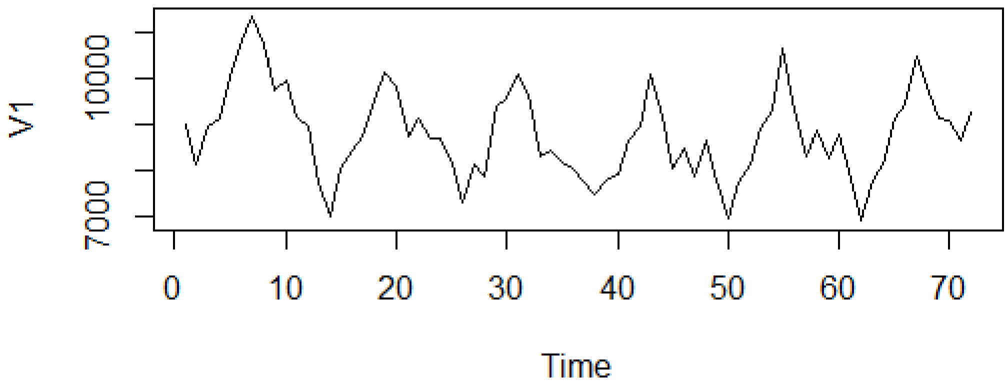
P/ACF large at lag 12; ⇒ think SARIMA P=1 or Q=1 or P=Q=1

ACF ρ(1) ≠0, with {ρ(k), k=2, …, 11}, within confidence intervals ⇒ suspect MA, q=1

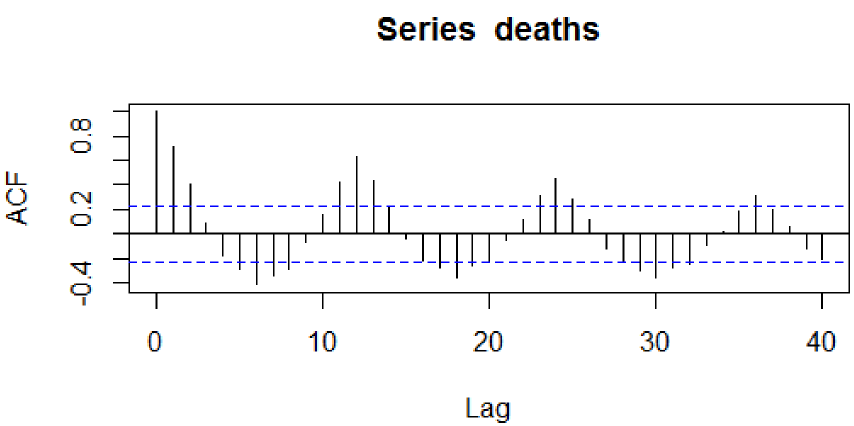
PACF α(1) ≠0, with {α (k), k=2, …, 11}, within confidence intervals ⇒ suspect AR, p=1

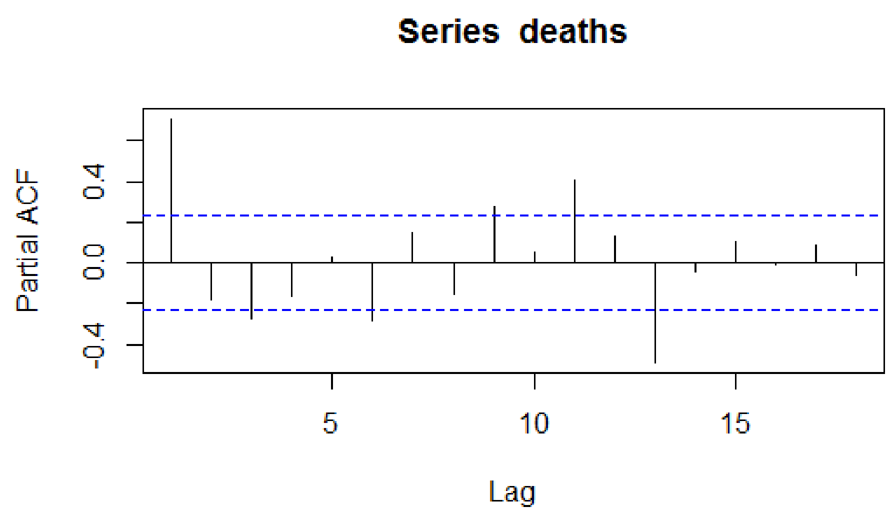
Also consider p=q=1.

ACF for Accidental Deaths Data
—what happens if seasonality is not removed

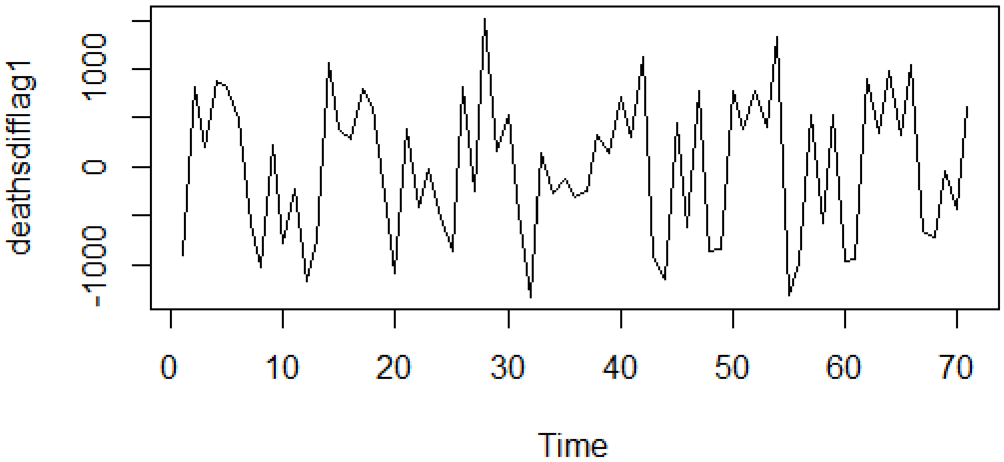


Seasonality not removed;
acf periodic,
remains large for large lags

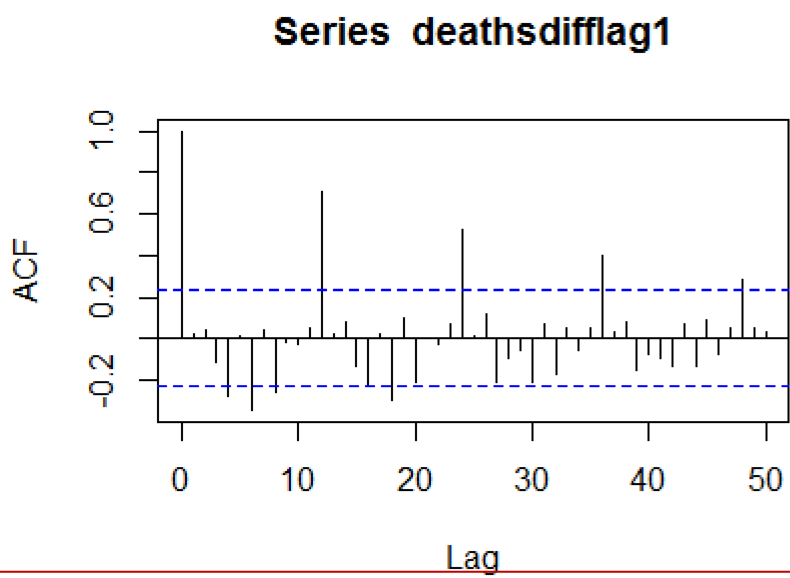


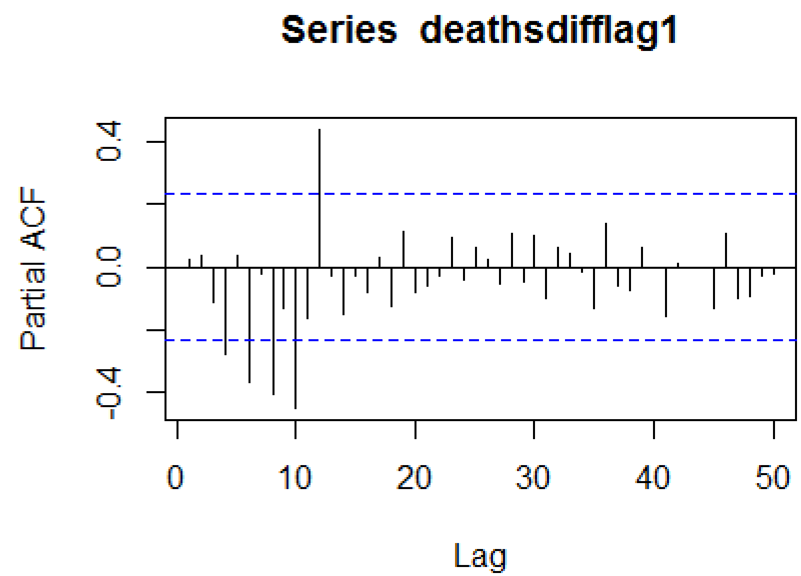


ACF for Accidental Deaths Data, Differenced at lag 1



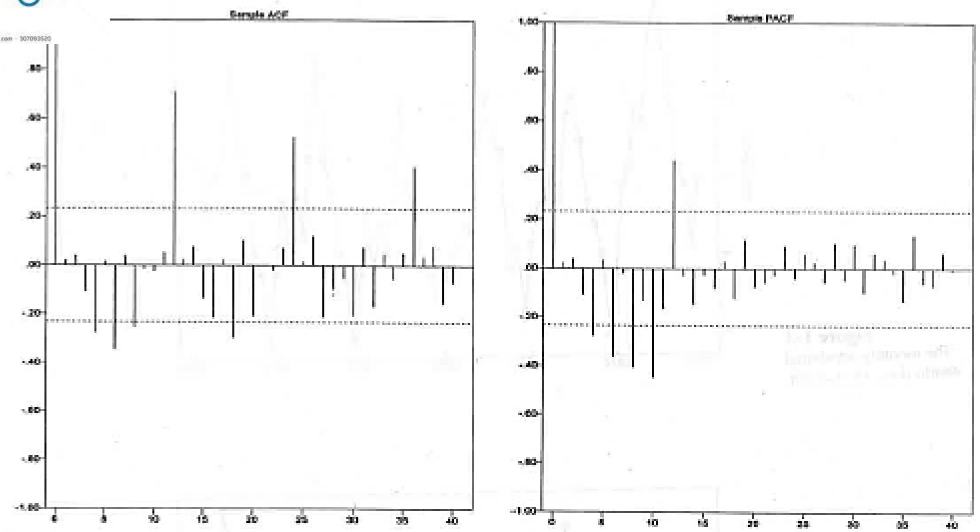
Seasonality not removed; acf periodic.



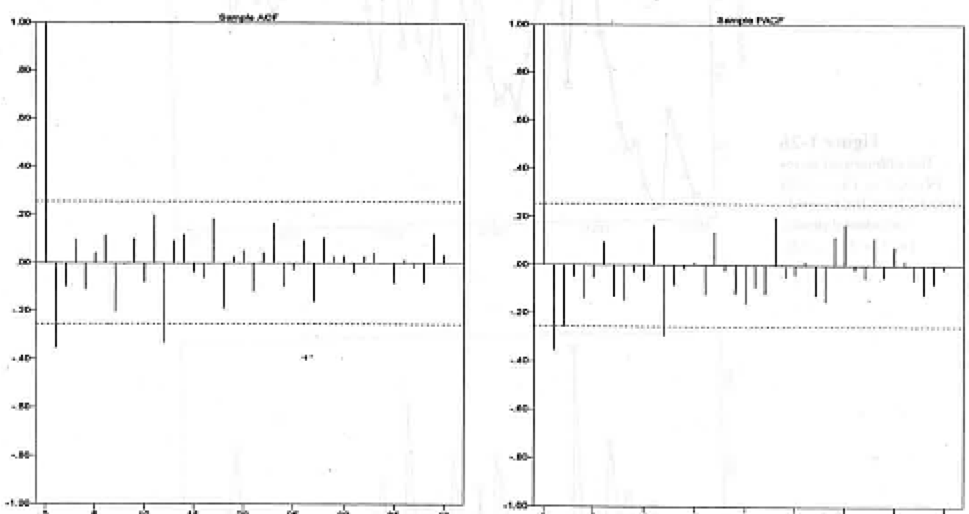


Much harder to choose a model; ACF remains large.
PACF suggests SAR at s = 12, P=1, but monthly dependence p, q hard to determine

ACF for Accidental Deaths Data and its differences



$\nabla(\text{Deaths})$



$\nabla\left(\nabla_{12}(\text{Deaths})\right)$

Note the difference:

Row 1: Seasonality not removed;

acf periodic and large

Row 2: Seasonality removed;

acf die out

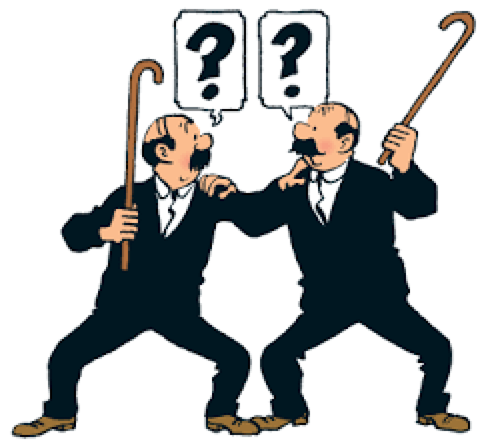
Check your understanding

Consider the model $(1 - B)(1+0.8B)X_t = (1- 0.5B^4)\, Z_t,\ Z_t \sim WN(0, \sigma_Z^2)$

(a) Identify the model as one of the ARMA, ARIMA, SARIMA, specify all parameters

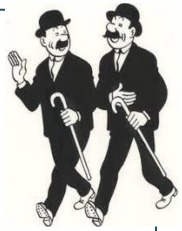
(b) Verify whether the model is stationary or invertible or both.

(c) If you believe that the model is non-stationary, what should a modeler do to make it stationary?



check your answers on the next slide

Check your understanding



Consider the model $(1 - B)(1+0.8B)X_t = (1- 0.5B^4) Z_{t,}$, $Z_t \sim WN(0, \sigma_Z^2)$

(a) Identify the model as one of the ARMA, ARIMA, SARIMA, specify all parameters

Solution 1: This is SARIMA$(1, 1,0)(0,0,1)_4$ model :

-- MA part has one term $(1- 0.5B^4) Z_t$, thus Q=1, s=4, and q=0.

-- Term (1-B) in the l.h.s. (left-hand side) corresponds to differencing at lag 1 once, i.e. d=1. No differencing at lag 4, so D=0.

-- AR part has term $(1+0.8B)X_t$, thus, p=1, P=0. (No AR term with B^4)

Solution 2: This is ARIMA(1, 1, 4) model:

-- MA part has terms $Z_t – 0.5 Z_{t-4}$, corresponding to MA(4) with $\theta_k = 0$ for k=1,2,3.

-- Term (1-B) in the l.h.s. corresponds to differencing at lag 1 once, i.e. d=1.

-- AR part has term $(1+0.8B)X_t$, thus, p=1.

Both solutions are correct!

(b) Verify whether the model is stationary or invertible or both.

Because $|\Theta_1| = 0.5 < 1$, the model is invertible. (For solution 2, check roots of $\theta(z) = 1 – 0.5z^4$.)

Because AR part has a unit root, it is not stationary.

(c) If you believe that the model is non-stationary, what should a modeler do to make it stationary?

The modeler should difference the data once at lag 1: $Y_t = (1- B)X_t$, to get stationary and invertible ARMA(1, 4) model $(1+0.8B)Y_t = (1- 0.5B^4) Z_{t,}$, $Z_t \sim WN(0, \sigma_Z^2)$

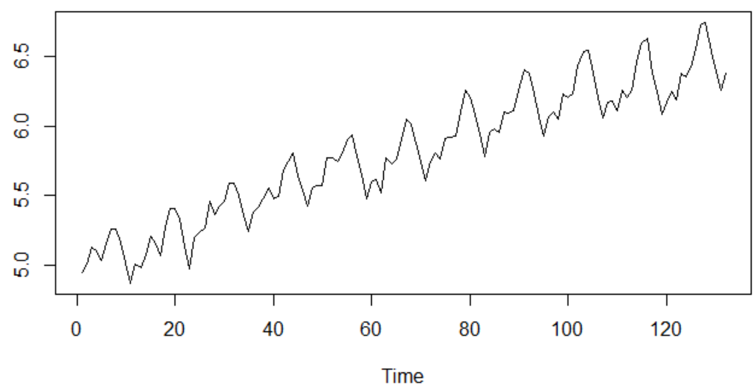
Check your understanding

(a) Write an equation corresponding to SARIMA (1, 1, 0) x (1, 1, 2)$_6$

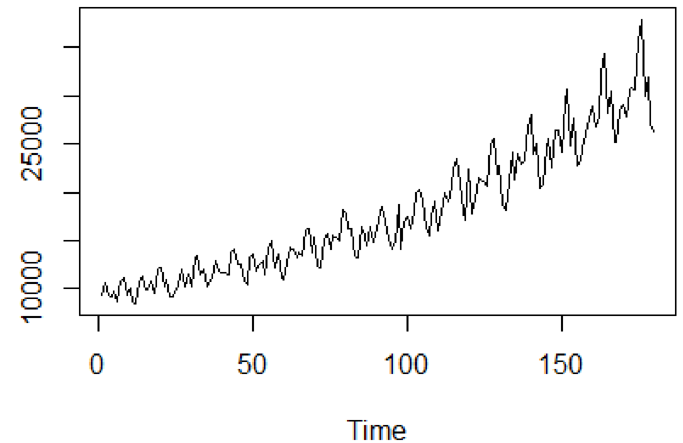
(b) Would SARIMA model appropriate for datasets with the plots below?

If you believe that it is not appropriate, what is your advice to the modeler?

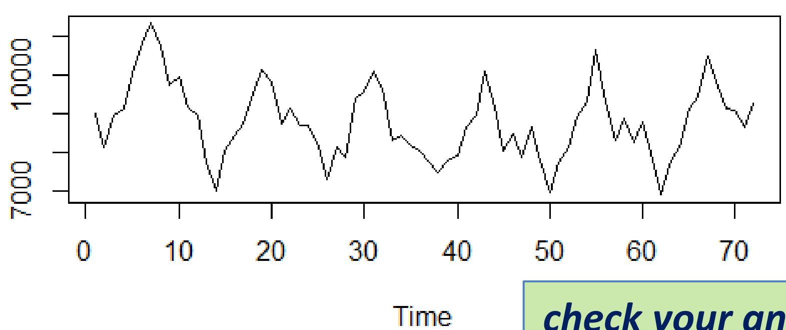
A.

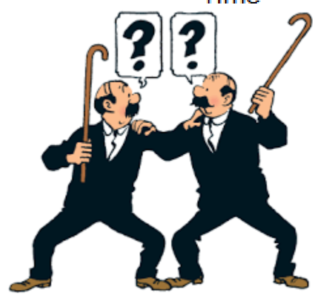


B.



C.





check your answers on the next slide

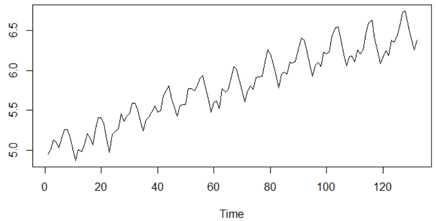
Check your understanding

(a) Write an equation corresponding to SARIMA (1, 1, 0) x $(1, 1, 2)_6$:

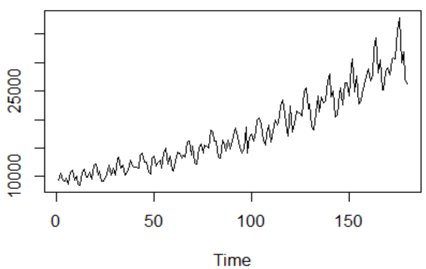
Identify: (1,1,0) ⇒ (p=1, d=1, q=0); $(1, 1, 2)_6$ ⇒ (P=1, D=1, Q=2, s=6).

Model: $(1 - \Phi_1 B^6)(1 - \phi_1 B)(1 - B^6)(1 - B)\, X_t = (1 + \Theta_1 B^6 + \Theta_2 B^{12})\, Z_t$, $Z_t \sim WN(0, \sigma_Z^2)$

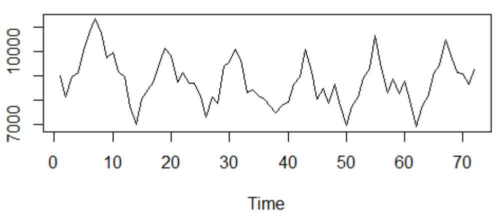
A.



B.



C.



(b) Would SARIMA model appropriate for datasets with the following plots: A,B,C.
If you believe that it is not appropriate, what is your advice to the modeler?

SARIMA is not appropriate for B. because this dataset has an increasing variance.
The modeler should apply an appropriate transformation to stabilize variance and then consider a SARIMA model.

SARIMA model seems to be appropriate for A & C.

Check your understanding

Consider SARIMA $(0,0,1)(0,0,1)_{12}$ model.

Determine all positive lags k at which the autocorrelation is non-zero.

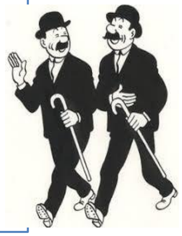
A. 1
B. 1, 12
C. 1, 11, 12
D. 1, 11, 12, 13
E. 1, 2, 11, 12, 13

check your answers on the next slide

Check your understanding

Consider SARIMA $(0,0,1)(0,0,1)_{12}$ model.
Determine all positive lags k at which the autocorrelation is non-zero.

A. 1
B. 1, 12
C. 1, 11, 12
D. 1, 11, 12, 13
E. 1, 2, 11, 12, 13



Short answer: D.

Long answer: The model (p=P=0, q=Q=1) was considered in Example 9.3 of Part III.
We saw that it has non-zero ACFs at lags 1, 11, 12, and 13.
That is, Q=1 gives $\rho_X(12) \neq 0$; in addition, q=1 gives $\rho_X(1) \neq 0$
and also $\rho_X(11)$ & $\rho_X(13) \neq 0$ (lag 1 before and after the lag s=12).

Even longer answer:

The model can be written as MA(13):

$$X_t = (1 + \theta_1 B)(1 + \Theta_1 B^{12}) Z_t = Z_t + \theta_1 Z_{t-1} + \Theta_1 Z_{t-12} + \Theta_1 \theta_1 Z_{t-13}\,.$$

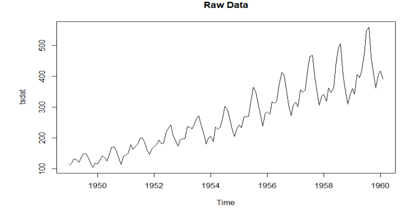
This is a special case of MA(13) with only few non-zero coefficients: $\theta_k = 0;\ k \neq 1, 12, 13$.

ACF of MA were found in Week 2, slides part III of Lecture 3 and § 3.4 of Lecture Notes:

$$\rho_X(k) = \frac{\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \ldots + \theta_{q-k}\theta_q}{1 + \theta_1^{\,2} + \theta_2^{\,2} + \ldots + \theta_q^{\,2}}; \quad \rho_X(k) = 0,\ k > q.$$

For general θ_1 & Θ_1, the numerator is not zero
for k=1, 12, 13 because of the non-zero first term θ_k;
for k=11 because of the second term $\theta_1\theta_{12}$.

Main Points of Lecture 7: SARIMA



SARIMA (p,d,q) x (P, D, Q)$_s$

- The model is for a data set with stable variance. If necessary, transform data.

A. Difference D times at lag s to remove seasonality.

B. Difference d times at lag 1 to remove trend.

- At this point the process should be stationary.
- View time series X_t as s series: X_j, X_{j+s}, …, $X_{j+(r-1)s}$, j=1, 2, …,
  -- for January: j = 1; for February: j=2; …, ; for December : j=12=s.—

C. Consider s datasets, each containing of annual entries for the same month j
It is a stationary time series. Model it by ARMA(P,Q): $\Phi(B^s)Y_t = \Theta(B^s)U_t$

- Assumption: annual data for each month, is modelled by the same ARMA(P,Q) .
- To identify P & Q, consider sample P/ACF at lags sk, k = 1, 2, …

D. Consider a dataset of s entries within one period. This is a stationary dataset.
Model it by ARMA(p,q): $\phi(B)U_t = \theta(B)Z_t$

- Assumption: for each year, 1980, 1981, etc., within each year, data follows ARMA(p,q)
- To identify p & q, consider sample P/ACF at lags k = 1, 2 , …, s-1.

E. Final SARIMA (p,d,q)(P,D,Q)$_s$ Model

$$\Phi(B^s)\phi(B)(1-B)^d(1-B^s)^D X_t = \Theta(B^s)\theta(B)Z_t,\ Z_t \sim WN(0,\sigma_Z^2).$$

REMEMBER:

$$\text{SARIMA}\ \underbrace{(p,d,q)}_{\substack{\uparrow \\ \left(\begin{array}{l}\text{Non-seasonal part}\\ \text{of the model}\end{array}\right)}}\ \underbrace{(P,D,Q)_m}_{\substack{\uparrow \\ \left(\begin{array}{l}\text{Seasonal part}\\ \text{of the model}\end{array}\right)}}$$

SARIMA

Southern African Research & Innovation Management Association

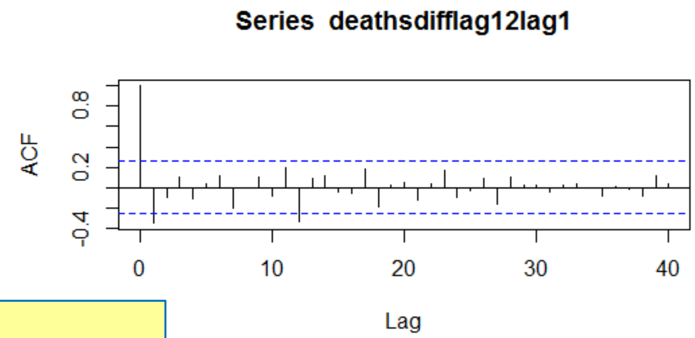
Leadership in Research & Innovation Management

Sample ACF/PACF Estimation. Confidence Intervals

… *Sample correlation is almost zero…*

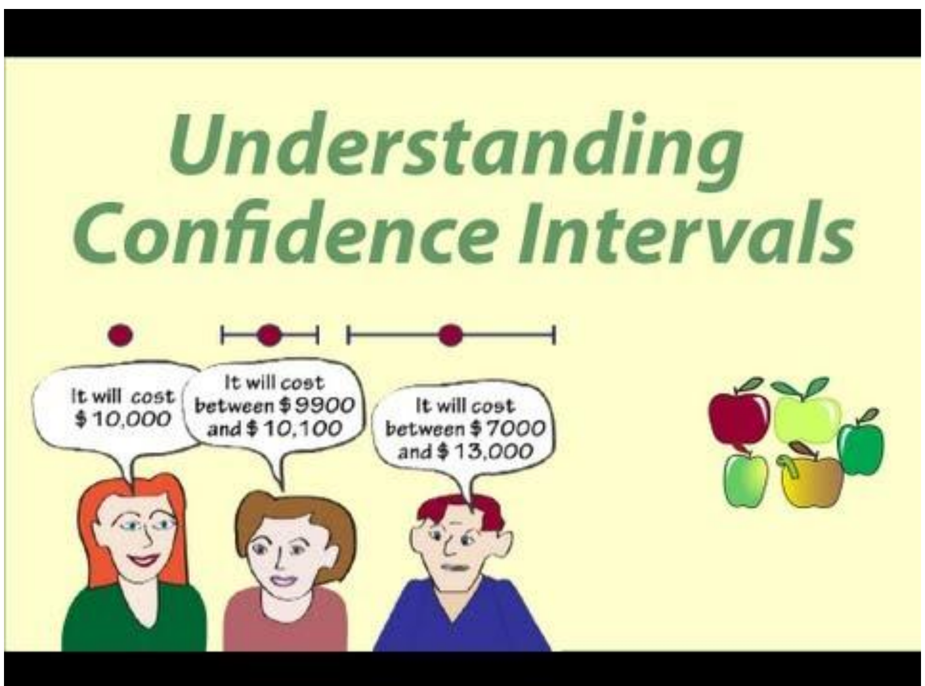
What does this statement mean?

What is small?

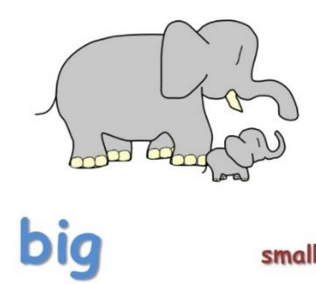


Today's Question # 2:
How do we decide that sample ACF/PACF can be taken as zero?

What are dashed blue lines on the graphs?



To answer: need to understand confidence intervals



Connection with previous material:

Models:

(Kazimir Malevich)

Lectures 1-7: Introduced Models:

Classical Decomposition Model: $X_t := m_t + s_t + S_t$
AR(p), MA(q), ARMA(p, q), ARIMA(p,d,q), SARIMA (p,d,q) x $(P,D,Q)_s$

Tools:

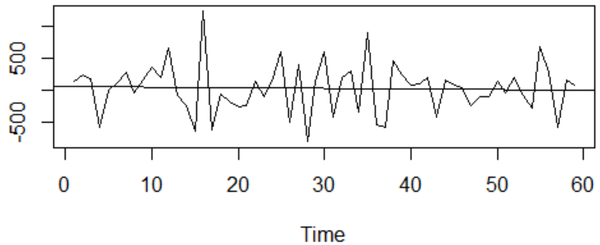
Lectures 1-7: Introduced Tools:

Tools to remove non-stationary terms m_t and s_t : $\nabla^d X_t \, , \nabla_s{}^D X_t$

Tools of Model Identification: ACF, PACF

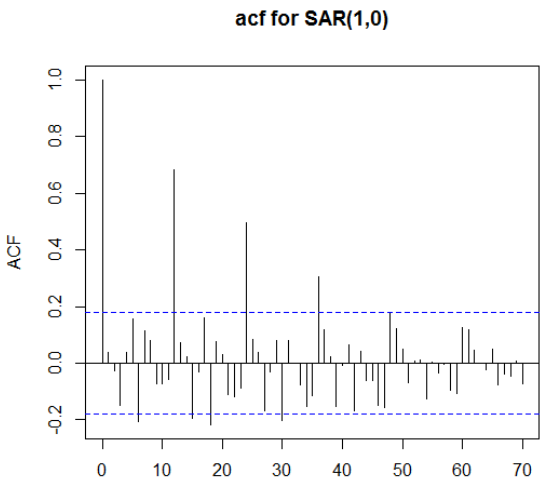
(Vasily Tropinin)

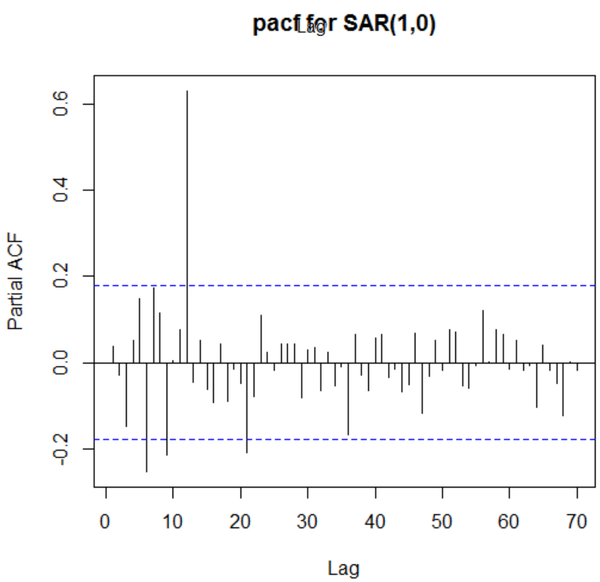
Today's Topic: how to estimate ACF/PACF from Data

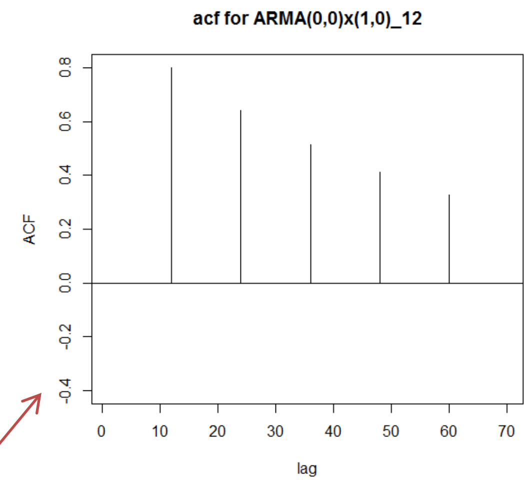


*'Data! Data! Data!' he cried impatiently. '
I can't make bricks without clay.'"*

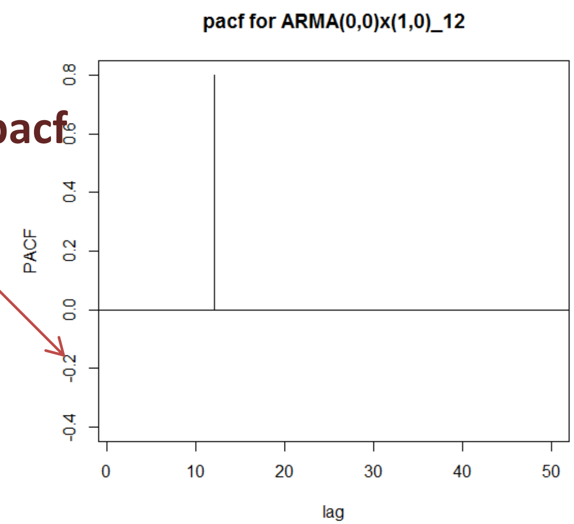
Example: Sample acf/pacf for the simulated model: $X_t - 0.8\,X_{t-12} = Z_t$







theoretical acf/pacf



Confidence intervals help to decide whether sample acf/pacf at particular lag is significant

10. Estimation of Sample ACF/PACF.

Question 1: How the graphs of Sample ACF/PACF created? What does the R command do?

Let's start with time series data:

From WEEK 1:

- **Time Series Data: a series of values recorded in time**
- **Time-Ordered & Equally Spaced Dependent Observations**
- **Notation: $X_1, X_2, X_3, \dots, X_{t, \dots,}$**

From WEEK 3:

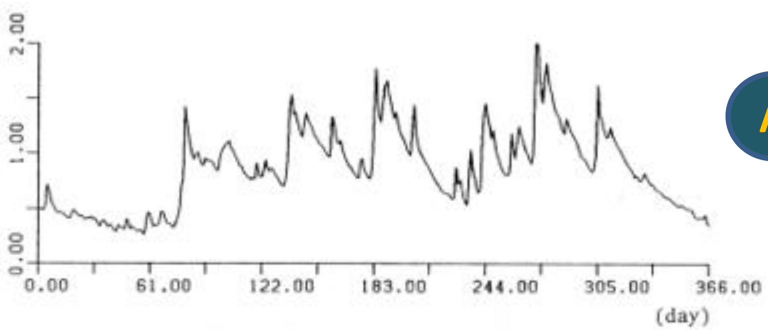
1. **Tools ACF/PACF are developed for stationary series X;**
2. **Box-Jenkins Methodology works for nonstationary time series with polynomial trend and period: $X_t = m_t + s_t + S_t$**
3. **Use differencing to remove non-stationarity and apply tools from (1).**
4. **Also might work if a transformation applies.**

10. Estimation of Sample ACF/PACF.

START with EXAMINING DATA

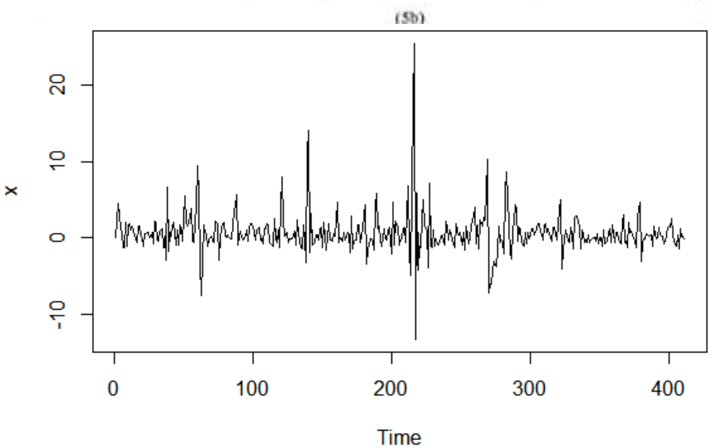
We learned to handle the following types of non-stationarity:

- $X_t = m_t + s_t + S_t$ = trend + seasonality + stationary time series
- Possible to transform to make data closer to stationary Gaussian

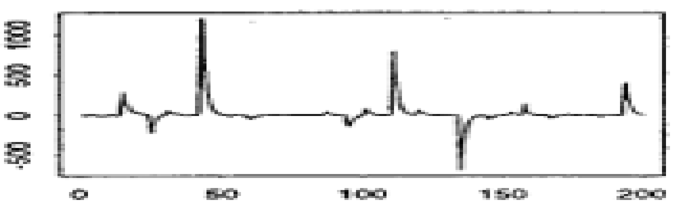


A

Choose your datasets carefully; examine data before starting modelling



B



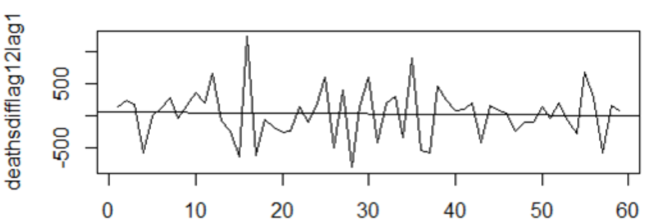
C

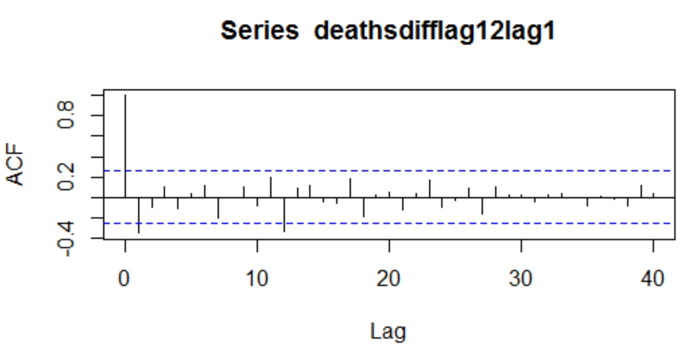
Can our current methods handle these data sets?

A: These datasets do not look like datasets we saw in the class

Week 4 Lecture 8 Part II: Working with Data. Sample ACF and its Distribution

Outline of Part II of Lecture 8:





How the graphs of Sample ACF created?
What does the R command do?

To answer: need formulas

10.1 Estimation of Sample Mean.

Question 1: How the graphs of Sample ACF/PACF created? What does the R command do?

Step 1: Plot Data. Make your data STATIONARY (transform, differencing)

Step 2: Use stationary sample $X_1, X_2, X_3, \dots, X_n$ to estimate moments

10.1 Sample Mean

- **Formula:** $\bar{X}_n = \frac{1}{n}\sum_{t=1}^{n} X_t$;
- **Stationarity required: must have $E(X_t) = \mu_X$ be the same for all t;**
- **Unbiased:** $E(\bar{X}_n) = \frac{1}{n}\sum_{t=1}^{n} E(X_t) = \frac{1}{n}\sum_{t=1}^{n} \mu_X = \mu_X$.
- **Confidence Interval:**

For large sample, $\bar{X}_n$ is approximately normal (see [BD], § 2.4.1)

$\bar{X}_n \approx N(\mu_X, v/n)$ with $v \approx \gamma_X(0) + 2\sum_{k=1}^{\infty} \gamma_X(k)$.

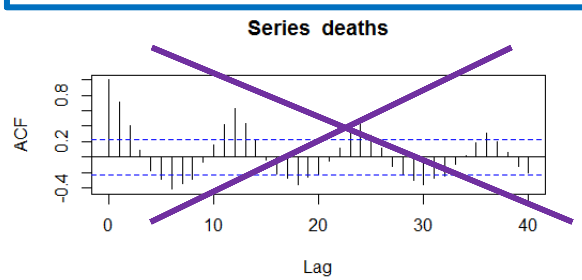
95% C.I. for μ_X : $(\bar{X}_n - 1.96\sqrt{v}/\sqrt{n}, \bar{X}_n + 1.96\sqrt{v}/\sqrt{n})$

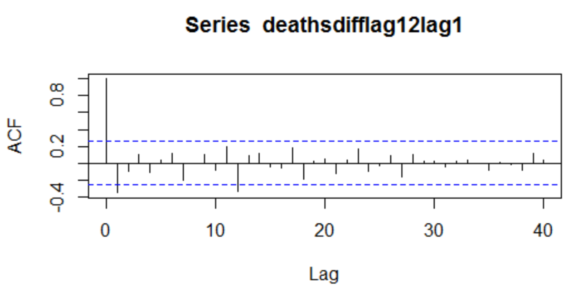
In practice, v is replaced by $\hat{v}$ -- see formula in lecture notes or [BD]

10.2 Estimation of Sample Variance and ACF

Question 1: How the graphs of Sample ACF/PACF created? What does the R command do?

- **Sample Variance:** $\hat{\sigma}_X{}^2 \equiv \hat{\gamma}_X(0) = \frac{1}{n}\sum_{t=1}^{n}(X_t - \bar{X}_n)^2$;
- **Sample ACVF at lag h:** $\hat{\gamma}_X(h) = \frac{1}{n}\sum_{t=1}^{n-h}(X_t - \bar{X}_n)(X_{t+h} - \bar{X}_n)$;
- **Sample ACF at lag h:** $\hat{\rho}_X(h) = \hat{\gamma}_X(h)/\hat{\gamma}_X(0)$;
- **n & h sizes:** $n \ge 50$; $h \le n/4$ **(Box & Jenkins);**
- **Stationarity required:** $\gamma_X(h)$ **is the same for all t;**
- **Characteristics of sample ACF patterns for non-stationary series:**
  - **For nonstationary TS usually** $|\hat{\rho}_X(h)|$ **stays large for large lags**
  - **For data with strong deterministic periodic component, sample ACF is periodic with the same period**





10.2 Estimation of Sample Variance and ACF

Question 1: How the graphs of Sample ACF/PACF created? What does the R command do?

- **Sample Variance:** $\hat{\sigma}_X \equiv \hat{\gamma}_X(0) = \frac{1}{n}\sum_{t=1}^{n}(X_t - \bar{X}_n)^2$;
- **Sample ACVF at lag h:** $\hat{\gamma}_X(h) = \frac{1}{n}\sum_{t=1}^{n-h}(X_t - \bar{X}_n)(X_{t+h} - \bar{X}_n)$;
- **Sample ACF at lag h:** $\hat{\rho}_X(h) = \hat{\gamma}_X(h)/\hat{\gamma}_X(0)$;

Question 2: What are dashed blue lines on the graphs?

Confidence Intervals based on Bartlett's Formula:

For large sample, $\hat{\rho}_X(h)$ are approximately normal with known variance:

$$\hat{\rho}_X(h) \approx N(\rho_X(h), w_{hh}/n)$$ (*see formula for w_{hh} on the next slide*)

This allows to write 95% confidence interval (*see examples on next slides*)

10.3 Estimation of Sample Variance and ACF

Question 2: What are dashed blue lines on the graphs?

10.3 Asymptotic Distribution of Sample ACF -- Bartlett's Formula

Assumptions:

(i) Process $X_1, X_2, X_3, \ldots, X_n$ is stationary with $Z_t \sim$ I.I.D. $(0, \sigma_Z^{\,2})$;

(ii) Sample size n is large

Box-Jenkins recommend $n \geq 50$ and lag $h \leq n/4$;

For this class recommend $n \geq 100$ or more.

(b/c 100/4 = 25, once cannot see seasonal effects for period 12)

Main Result:

Let $\underline{\rho}_h = (\rho(1), \rho(2), \ldots, \rho(h))'$ be vector of ACFs of X_t.

Let $\underline{\hat{\rho}}_h$ be the sample ACF. Then, vector of sample acfs , $\underline{\hat{\rho}}_h$, is approximately normal with known variance:

$$\underline{\hat{\rho}}_h \approx N(\underline{\rho}_h\, , \; n^{-1}W)\,;$$

Entries of the covariance matrix W:

$$w_{ij} = \sum_{k=1}^{\infty}\{\rho(k+i) + \rho(k-i) - 2\rho(i)\,\rho(k)\}\{\rho(k+j) + \rho(k-j) - 2\rho(j)\,\rho(k)\}$$

10.3.1 Example: Sample ACF of I.I.D. WN

Question 2: What are dashed blue lines on the graphs?

Bartlett's Formula :

Vector of Sample ACF $\underline{\hat{\rho}_h} \approx N(\underline{\rho}_h, n^{-1}W)$ with entries of W

$w_{ij} = \sum_{k=1}^{\infty}\{\rho(k+i) + \rho(k-i) - 2\rho(i)\,\rho(k)\}\{\rho(k+j) + \rho(k-j) - 2\rho(j)\,\rho(k)\}$

10.3.1 Application to I.I.D. WN

- For I.I.D. WN $\rho(k) = 0$ for all $k \neq 0$ and $\rho(0) = 1$. Calculate:
- $w_{ij} = 0$ if $i \neq j$; $w_{ii} = 1$ if $i = j$ *(the term $\rho(k-i) = \rho(k-j) = 1$ when $i = j = k$)*
- Thus, $\hat{\rho}_Z(1), \ldots, \hat{\rho}_Z(h)$ are approximately i.i.d $N(0, 1/n)$:
  - $\hat{\rho}_Z(h) \approx N(\rho_Z(h) = 0, w_{hh}/n = 1/n)$,
  - Uncorrelated Normal $\Rightarrow$ independent

More detailed explanation on the next slide

- Using 95% Confidence Interval, conclude:

If $|\hat{\rho}(h)| < 1.96/\sqrt{n}$ for all $h \geq 1$, assume WN = MA(0).

10.3.1 Example: Sample ACF of I.I.D. WN-cont'd

More Explanations on Confidence Intervals for WN:

10.3.1 Application to I.I.D. WN

For WN, we concluded that

- **Thus, $\hat{\rho}_Z(1), \ldots, \hat{\rho}_Z(h)$ are approximately i.i.d N(0, 1/n):**
  - **$\hat{\rho}_Z(h) \approx N(\rho_Z(h)=0, w_{hh}/n = 1/n)$,**
  - **Uncorrelated Normal $\Rightarrow$ independent**
- **This means that with probability of .95,** $\frac{|\hat{\rho}(h) - 0|}{\sqrt{1/n}} < 1.96$
- **This means that if we plot sample autocorrelation function as a function of lags h, approximately 95% of the sample acfs should lie between bounds** $\pm 1.96 n^{-1/2}$.
- **Conclude: with 95% confidence,**

$\boxed{\text{If } |\hat{\rho}(h)| < 1.96/\sqrt{n} \text{ for all } h \geq 1, \text{ assume WN = MA(0).}}$

10.3.2 Example: Sample ACF of MA(q)

Question 2: What are dashed blue lines on the graphs?

Bartlett's Formula :

Vector of Sample ACF $\underline{\hat{\rho}_h} \approx N(\underline{\rho}_h, n^{-1}W)$ **with entries of W**

$$w_{ij} = \sum_{k=1}^{\infty} \{ \rho(k+i) + \rho(k-i) - 2\rho(i)\,\rho(k)\} \{\rho(k+j) + \rho(k-j) - 2\rho(j)\,\rho(k)\}$$

10.3.2 Application to MA(q) with IID WN:

$X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \cdots + \theta_q Z_{t-q}$ where $Z_t \sim$ I.I.D. $(0, \sigma_Z^2)$

- **From Bartlett's Formula,** $\hat{\rho}_X(h) \approx N(\rho_X(h), n^{-1} w_{hh})$
- **For MA(q),** $\rho(q+j) = 0$ **for all** $j > 0$**, so that**
  - ***For*** $h > q$ $\hat{\rho}_X(h) \approx N(0, n^{-1} w_{hh})$ ***with*** *(some algebra)*
  - $w_{hh} = 1 + 2\sum_{k=1}^{q} \rho^2(k)$
- **Using 95% Confidence Interval, conclude:**

  If $|\hat{\rho}_X(h)| < 1.96/\sqrt{n}$ **for all** $h \geq 1$**, assume MA(q) with q=0, that is, WN.**

10.3.2 Example: Sample ACF of MA(q)

Question 2: What are dashed blue lines on the graphs?

10.3.2 Application to MA(q) with IID WN:

$X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \cdots + \theta_q Z_{t-q}$ where $Z_t \sim$ I.I.D. $(0, \sigma_Z^2)$

- For $h > q$, $\hat{\rho}_X(h) \approx N(0, n^{-1} w_{hh})$
- $w_{hh} = 1 + 2\sum_{k=1}^{q} \rho^2(k)$

Identification of order q of MA(q) using 95% Confidence Intervals:

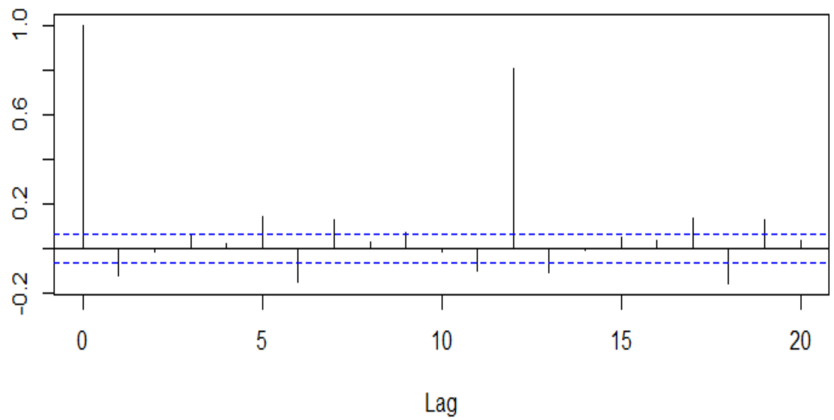
- R replaced the term $\{2 n^{-1} \sum_{k=1}^{q} \rho^2(k)\}$ in $n^{-1} w_{hh}$ by 0: $n^{-1} w_{hh} \approx \frac{1}{n}$;
- If $|\hat{\rho}_X(q_0)| > 1.96/\sqrt{n}$ but $|\hat{\rho}_X(h)| < 1.96/\sqrt{n}$ for all $h > q_0$,

  then assume MA(q) with $q = q_0$.
- Because R omits (unknown) term $2 n^{-1} \sum_{k=1}^{q} \rho^2(k)$ *when plotting C.I.,*

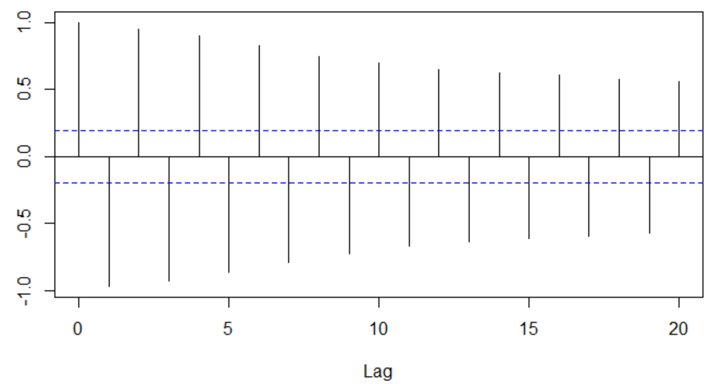
in case when $\hat{\rho}_X(h) \approx 1.96/\sqrt{n}$, assume that $\hat{\rho}_X(h)$ lies within the C.I.

Check your understanding

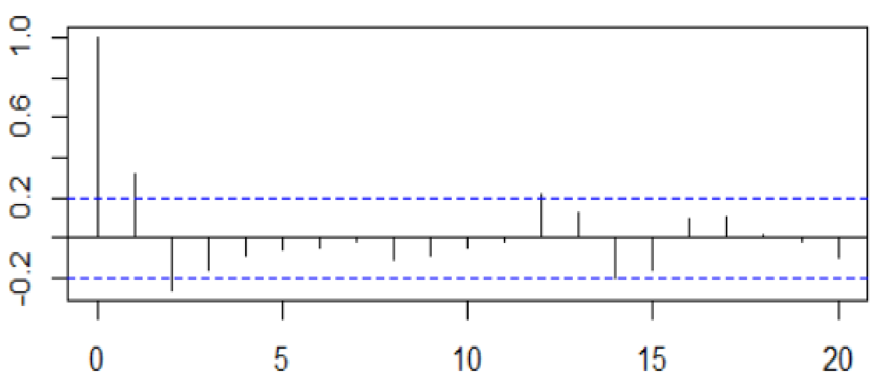
A



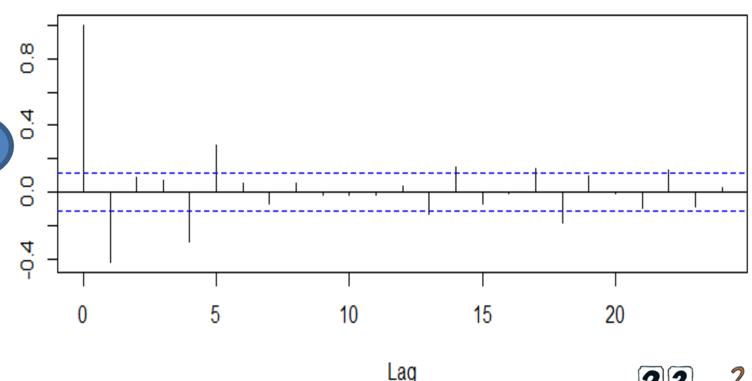
C



B



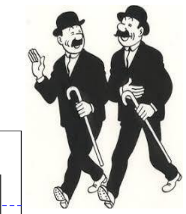
D



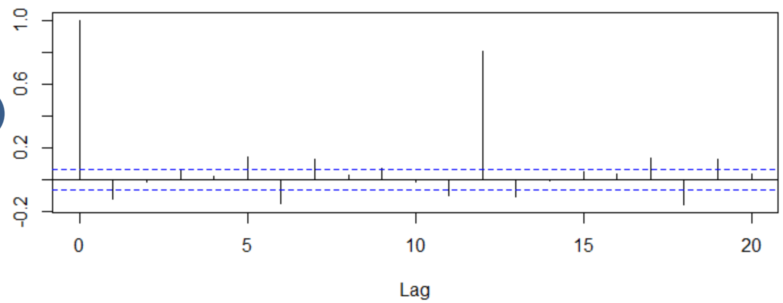
Question 2: What are dashed blue lines on the graphs?

Identify graphs of sample ACFs corresponding to MA(q) processes. Determine q.

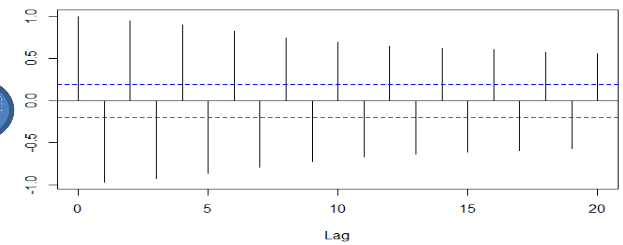
Check your understanding



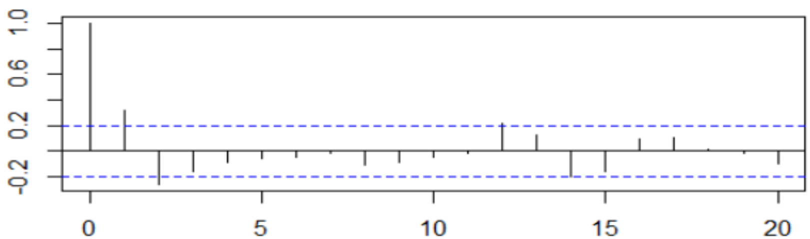
A



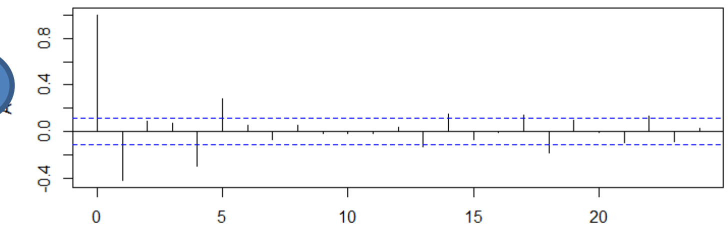
C



B



D



Identify graphs of sample ACFs corresponding to MA(q) processes. Determine q.

Discussion:

(A) Sample ACFs at lags 1, (5, 6, 7), (11, 12, 13), (17, 18, 19) are outside the confidence interval. It might be MA(19) but most likely is a SARIMA with s=6 and Q of at least 3. Also q=1.

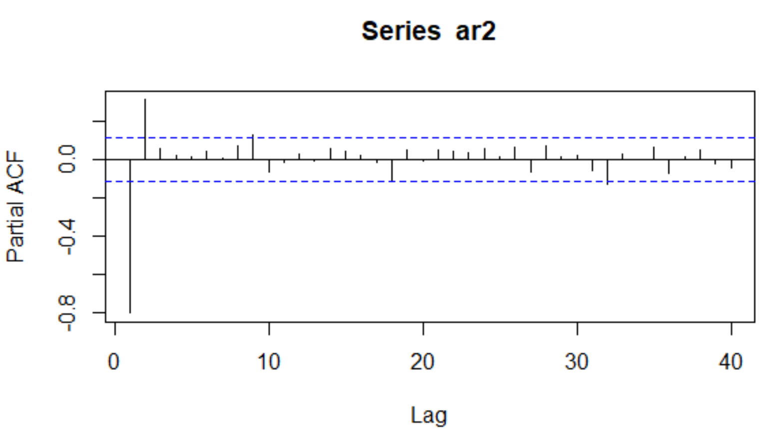
(B) Sample ACFs at lag 2 is outside the confidence interval, but is inside the intervals for lags k >2. Thus, the model is most likely MA(2).

(C) Sample ACF follows AR(p) pattern.

(D) Sample ACFs at lag 5 is outside the confidence interval, but is inside the intervals for lags k >5. Thus, the model is most likely MA(5).
(Acf at lag 18 is close to the border; disregard, recall Bartlett's formula.)

Week 4 Lecture 8: Working with Data: Sample PACF. Yule-Walker Parameter Estimation. The Durbin-Levinson Algorithm.

Outline of Part III of Lecture 8:



Questions:
How the graphs of Sample PACF created?
What does the R command do?
What the blue lines one the graph?

To answer: need formulas

10.4 Sample PACF: definition and distribution

Week 3 part I Lecture 6

Definition and

Examples of PACF

Yule-Walker equations in matrix form: $R_n\, \underline{\phi}_n = \underline{\rho}_n$;

Solve for n-dimensional vector $\underline{\phi}_n = R_n^{-1}\, \underline{\rho}_n$

PACF at lag n is $\alpha(n) = \phi_{nn}$**, the last component of** $\underline{\phi}_n$

Definition of

Sample PACF

Sample PACF: In the above definition, replace unknown acfs in matrix R_n **and vector** $\underline{\rho}_n$ **with sample acfs calculated from data**

From Slides Part I of Lecture 6 Week 3

$$\underline{\rho}_n = \begin{pmatrix} \rho_X(1) \\ \rho_X(2) \\ \cdot \\ \cdot \\ \rho_X(n) \end{pmatrix}, \quad \underline{\phi}_n = \begin{pmatrix} \phi_{n1} \\ \phi_{n2} \\ \cdot \\ \cdot \\ \phi_{nn} \end{pmatrix}, \quad R_n = \begin{pmatrix} 1 & \rho_X(1) & \dots & \rho_X(n-1) \\ \rho_X(1) & 1 & \dots & \rho_X(n-2) \\ \dots & \dots & \dots & \dots \\ \rho_X(n-1) & \rho_X(n-2) & \dots & 1 \end{pmatrix}$$

Then, Yule-Walker equations are: $R_n \underline{\phi}_n = \underline{\rho}_n$ and $\underline{\phi}_n = R_n^{-1} \underline{\rho}_n$.

The last component of $\underline{\phi}_n = R_n^{-1} \underline{\rho}_n$ is

$\alpha(n) \stackrel{def}{=} \phi_{nn}$ – PACF at lag n.

Sample PACF: replace acf with sample acf

10.4 Sample PACF: definition and distribution

Week 3 Part I Lecture 6
Definition and
Examples of PACF

Yule-Walker equations in matrix form: $R_n \underline{\phi}_n = \underline{\rho}_n$;
Solve for n-dimensional vector $\underline{\phi}_n = R_n^{-1} \underline{\rho}_n$
PACF at lag n is $\alpha(n) = \phi_{nn}$, the last component of $\underline{\phi}_n$

Definition of
Sample PACF

Sample PACF: replace unknown acfs in matrix R_n and vector $\underline{\rho}_n$ with sample acfs calculated from data

Distribution of sample PACF with application to identify AR(p) models:

- **For AR(p), PACF $\alpha(p+j) = 0$ for all $j > 0$, so that**

For large sample size, for AR(p), the sample PACFs at lags h > p, are approximately independent normal $\hat{\alpha}(h) \approx N(0, 1/n)$

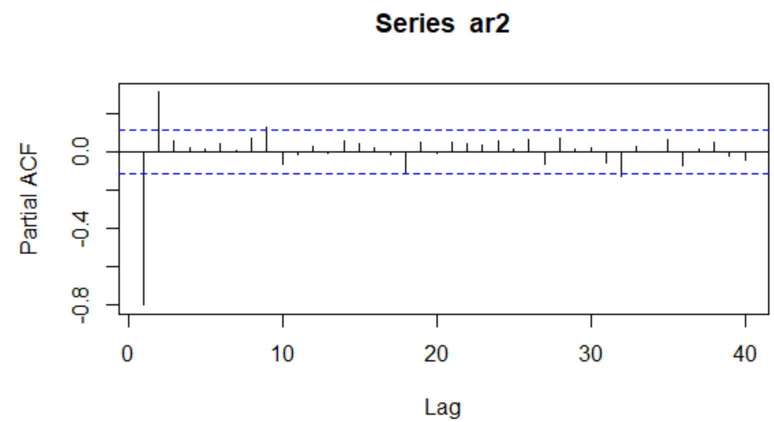
Identification of AR(p) model from stationary data:

If $|\hat{\alpha}(p_0)| > 1.96/\sqrt{n}$ & $|\hat{\alpha}(h)| < 1.96/\sqrt{n}$ for all $h > p_0$, assume AR(p_0)

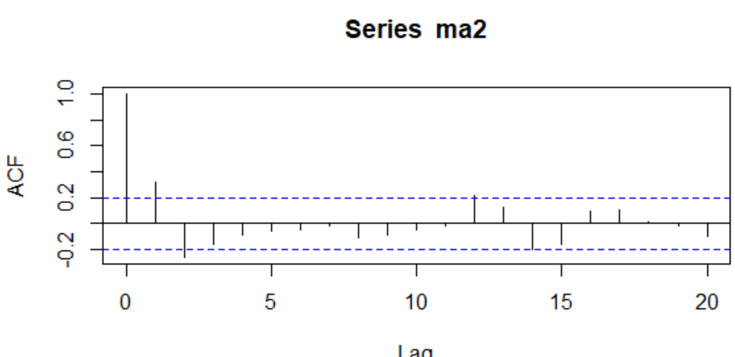
Summary: Identification of Parameters p in AR(p) and q in MA(q)

If sample PACF $\hat{\alpha}(n) \equiv \hat{\phi}_{nn} = 0$ for all $n > p$, but $\hat{\alpha}(p) \neq 0$, then the data corresponds to AR process of order p.

PACF for AR(p)



If sample ACF $\hat{\rho}(n) = 0$ for all $n > q$, but $\hat{\rho}(q) \neq 0$, then the data corresponds to MA process of order q.



ACF for MA(q)

Check your Understanding: AR(p) and MA(q): Model Identification

You are given the following information:

- **X and Y are two stationary time series**
- **The graphs below are (i) ACF of X; (ii) PACF of Y.**
- **The dashed lines above and below zero indicate the range within which the ACF and PACF results are considered not significantly different from zero.**

Which of the following statements displays the model structure that best describes series X and series Y?

Choose one

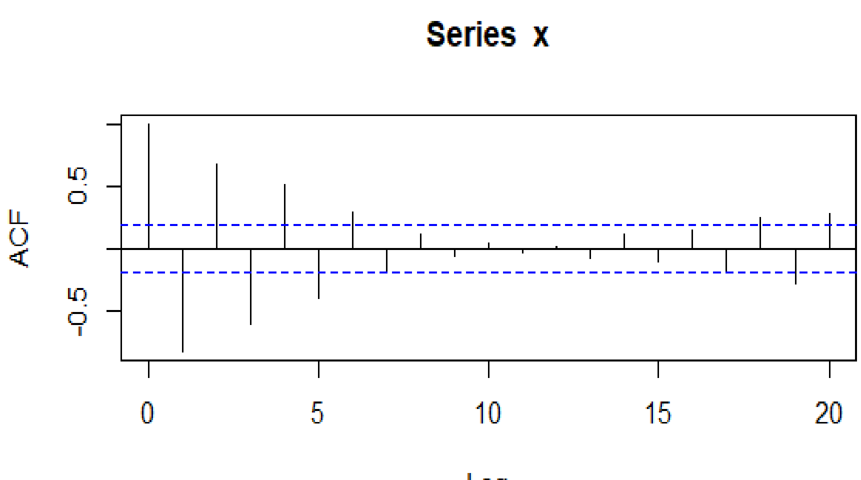
A. $X_t = 0.9\, X_{t-1} + Z_t$ **and** $Y_t = 0.9\, Y_{t-1} + Z_t + 0.6\, Z_{t-1} - 0.3 Z_{t-2}$

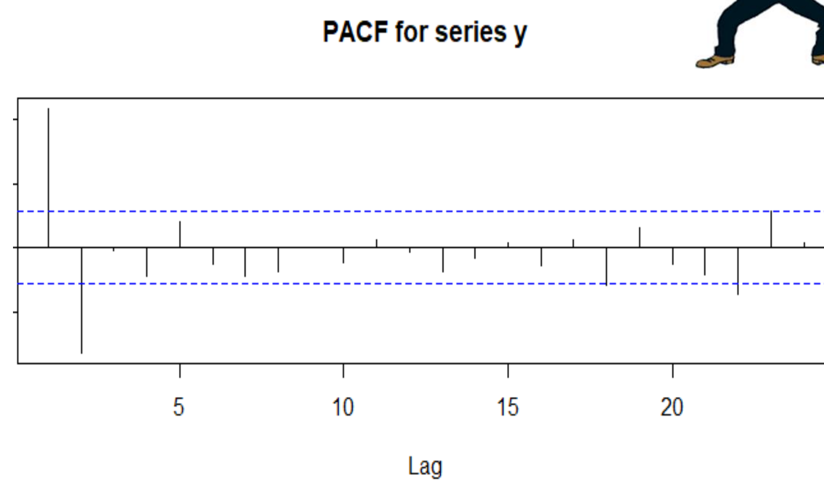
B. $X_t = -\,0.9\, X_{t-1} + Z_t$ **and** $Y_t = Z_t + 0.6\, Z_{t-1} - 0.3 Z_{t-2}$

C. $X_t = 0.9\, X_{t-1} + Z_t$ **and** $Y_t = Z_t + 0.6\, Z_{t-1}$

D. $X_t = -\,0.9\, X_{t-1} + Z_t$ **and** $Y_t = 0.6\, Y_{t-1} - 0.3 Y_{t-2} + Z_t$

E. $X_t = 0.9\, X_{t-1} + Z_t$ **and** $Y_t = 0.6\, Y_{t-1} - 0.3 Y_{t-2} + Z_t$





Check your Understanding: AR(p) and MA(q): Model Identification

You are given the following information:

- **X and Y are two stationary time series**
- **The graphs below are generated using the ACF functions of R**
- **The dashed lines above and below zero indicate the range within which the ACF results are considered not significantly different from zero.**

Which of the following statements displays the model structure that best describes series X and series Y?

A. $X_t = 0.9\,X_{t-1} + Z_t$ **and** $Y_t = 0.9\,Y_{t-1} + Z_t + 0.6\,Z_{t-1} - 0.3Z_{t-2}$

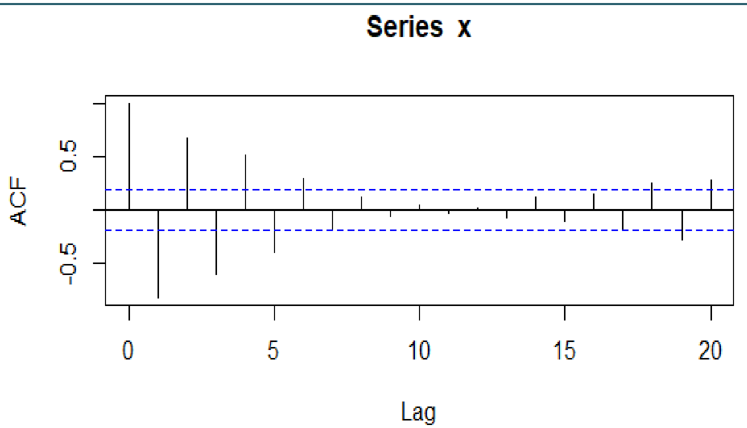
B. $X_t = -\,0.9\,X_{t-1} + Z_t$ **and** $Y_t = Z_t + 0.6\,Z_{t-1} - 0.3Z_{t-2}$

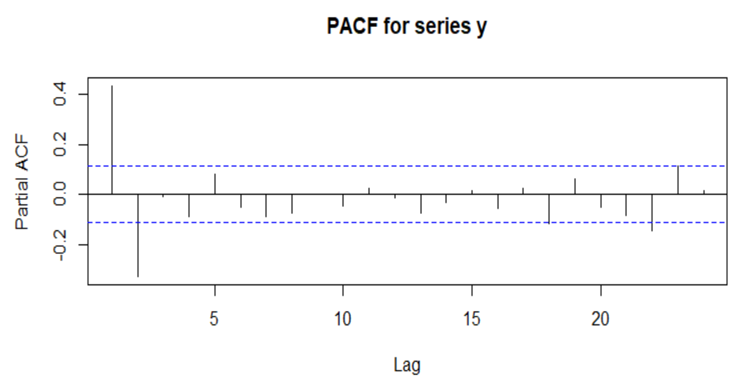
C. $X_t = 0.9\,X_{t-1} + Z_t$ **and** $Y_t = Z_t + 0.6\,Z_{t-1}$

D. $X_t = -\,0.9\,X_{t-1} + Z_t$ **and** $Y_t = 0.6\,Y_{t-1} - 0.3Y_{t-2} + Z_t$

E. $X_t = 0.9\,X_{t-1} + Z_t$ **and** $Y_t = 0.6\,Y_{t-1} - 0.3Y_{t-2} + + Z_t$

Choose one





(i) Identify X as AR(1) so that acf $\rho_X(k) = \phi_1^k$; (ii) From ACF of X, $\phi_1 < 0 \Rightarrow$ (B) or (D)
(iii) From PACF of Y, Y is AR(2), because for Y, $\phi_{kk} \approx 0$ for $k > 2$, $\phi_{22} \neq 0$.
$\Rightarrow$ **Answer: (D).**

10.5 Calculation of Sample PACF: the Durbin-Levinson Algorithm

- **Definition of Sample PACF: solve n x n system of linear equations $\hat{R}_n \, \underline{\widehat{\phi_n}} = \underline{\widehat{\rho_n}}$**

  Solve for n-dimensional vector $\underline{\widehat{\phi_n}}$. Define $\hat{\alpha}$(n) = $\hat{\phi}_{nn}$, the last component of $\underline{\widehat{\phi_n}}$.

- **To plot sample PACF for lags h=1,2, …, K, one has to invert K matrices $\hat{R}_h$**

- **In practice, use The Durbin-Levinson algorithm – recursive:**

- **The algorithm fits AR models of increasing orders h = 1, 2, … :**

$$X_t = \phi_{h-1,1}X_{t-1} + \phi_{h-1,2}\, X_{t-2} + \ \ldots + \phi_{h-1,h-1}\, X_{t-(h-1)} + Z_t$$

$$X_t = \phi_{h,1}X_{t-1} + \phi_{h,2}\, X_{t-2} + \ \ldots + \phi_{h,h}\, X_{t-h} + Z_t$$

noticing relationship between $\phi_{h,j}$ and $\phi_{h-1,j}$

Durbin-Levinson Algorithm gives the updating recursive equations:

$$\hat{\phi}_{hh} = \frac{\hat{\rho}(h) - \sum_{j=1}^{h-1} \hat{\phi}_{h-1,j}\hat{\rho}(h-j)}{1 - \sum_{j=1}^{h-1} \hat{\phi}_{h-1,j}\hat{\rho}(j)},$$

No need to memorize these formulas!

$$\hat{\phi}_{h,j} = \hat{\phi}_{h-1,j} - \hat{\phi}_{hh}\hat{\phi}_{h-1,h-j},\ j = 1, \ldots, h-1.$$

10.5.1 Example: Durbin-Levinson Algorithm for h=1,2,3.

Durbin-Levinson Algorithm gives the updating recursive equations:

$$\hat{\phi}_{hh} = \frac{\hat{\rho}(h) - \sum_{j=1}^{h-1} \hat{\phi}_{h-1,j}\hat{\rho}(h-j)}{1 - \sum_{j=1}^{h-1} \hat{\phi}_{h-1,j}\hat{\rho}(j)},$$

$$\hat{\phi}_{h,j} = \hat{\phi}_{h-1,j} - \hat{\phi}_{hh}\hat{\phi}_{h-1,h-j},\ j = 1, \ldots, h-1.$$

10.5.1 Example. Assume that sample ACF $\hat{\rho}(h)$ were calculated previously.

h=1: $\hat{\phi}_{11} = \hat{\rho}(1);$

. . . .

h=2: $\hat{\phi}_{22} = (\hat{\rho}(2) - \hat{\phi}_{11}\hat{\rho}(1))/(1 - \hat{\phi}_{11}\hat{\rho}(1)) = (\hat{\rho}(2) - \hat{\rho}^2(1))/(1 - \hat{\rho}^2(1)),$

$\hat{\phi}_{21} = \hat{\phi}_{11} - \hat{\phi}_{22}\hat{\phi}_{11};$

. . . .

h=3: $\hat{\phi}_{33} = (\hat{\rho}(3) - \hat{\phi}_{21}\hat{\rho}(2) - \hat{\phi}_{22}\hat{\rho}(1))/(1 - \hat{\phi}_{21}\hat{\rho}(1) - \hat{\phi}_{22}\hat{\rho}(2))$

$\hat{\phi}_{31} = \hat{\phi}_{21} - \hat{\phi}_{33}\hat{\phi}_{22}, \quad \hat{\phi}_{32} = \hat{\phi}_{22} - \hat{\phi}_{33}\hat{\phi}_{21};$ etc.

. . . .

11. Parameter Estimation for ARMA(p,q) Model

DONE:

Step 1: Plot Data. Make your data STATIONARY (transform, differencing)

Step 2: Use stationary sample $X_1, X_2, X_3, \ldots, X_n$ to estimate ACF/PACF

Step 3: Examine P/ACF, determine p & q for ARMA(p,q) model

Step 4: Estimate model p+q+1 parameters: $\phi_1, \ldots, \phi_p; \theta_1, \ldots. \theta_q; \sigma_Z^2$.

$$X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}, \quad Z_t \sim WN(0, \sigma_Z^2)$$

We cover:

11.1. Preliminary Estimation:

11.1.1. Yule-Walker Estimation for AR(p) models – The Durbin-Levinson Algorithm;

11.1.2 Preliminary estimation of MA(q) and ARMA(p,q) models – The Innovation Algorithm

11.2 MLE (maximum likelihood)and LSE (least –squares) estimation;

11.3 AICC criterion for order selection (Akaike Information Criterion, corrected for bias)

11.1 Preliminary Parameter Estimation: Method of Moments

11.1.1. Yule-Walker Estimation for AR(p) models.

Goal: Estimate model p+1 parameters: $\phi_1, \dots, \phi_p$; σ_Z^2 for AR(p) model

$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t$, $Z_t \sim WN(0, \sigma_Z^2)$

Method:

- **use relationship between parameters ϕ_j and second moments -- acfs;**
- **acfs can be replaced by sample acfs, obtained from data;**
- **Yule-Walker equations for AR(p) provide such a relationship: (Week 3 Lecture 5)**

$$R_p \underline{\phi}_p = \underline{\rho}_p; \qquad \gamma_X(0) \equiv \sigma_X^2 = \frac{\sigma_Z^2}{1 - \phi_1 \rho_X(1) - \dots - \phi_p \rho_X(p)}$$

- **Yule-Walker estimates are solutions of these equations with sample acfs substituted for unknown acfs:**

$$\underline{\widehat{\phi}_p} = \widehat{R}_p^{-1} \underline{\widehat{\rho}_p}; \; \widehat{\sigma}_Z^2 = \widehat{\gamma}(0)\{1 - \underline{\widehat{\phi}_p}' \underline{\widehat{\rho}_p}\} = \widehat{\gamma}(0)\{1 - \underline{\widehat{\rho}_p}' \widehat{R}_p^{-1} \underline{\widehat{\rho}_p}\}$$

11.1.1. Yule-Walker Estimation for AR(p) models.

Goal: Estimate model p+1 parameters: $\phi_1, \dots, \phi_p$; σ_Z^2 for AR(p) model

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t \text{ , } Z_t \sim \text{WN}(0, \sigma_Z^2)$$

- **Yule-Walker estimates are solutions of these equations with sample acfs substituted for unknown acfs:**

$$\underline{\widehat{\phi}}_p = \widehat{R}_p^{-1} \underline{\widehat{\rho}}_p \text{ ; } \widehat{\sigma}_Z^2 = \widehat{\gamma}(0)\{1 - \underline{\widehat{\phi}}_p{}' \underline{\widehat{\rho}}_p\} = \widehat{\gamma}(0)\{1 - \underline{\widehat{\rho}}_p{}' \widehat{R}_p^{-1} \underline{\widehat{\rho}}_p\}$$

Distribution of Yule-Walker Estimates of $\phi_1, \dots, \phi_p$; σ_Z^2 for AR(p) model

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t \text{ , } Z_t \sim \text{WN}(0, \sigma_Z^2) :$$

$$\underline{\widehat{\phi}}_p \approx N(\underline{\phi}_p , n^{-1} \sigma_Z^2 \, \Gamma_p^{-1});$$

Here $\Gamma_p = \{\gamma(k-j)\}_{j, k = 1, \dots p}$ is the auto covariance matrix

11.1.1. Yule-Walker Estimation for AR(p) models: The Durbin-Levinson Algorithm.

- **Yule-Walker estimates are solutions of these equations with sample acfs substituted for unknown acfs:**

$$\underline{\widehat{\phi_p}} = \hat{R}_p^{-1}\,\underline{\widehat{\rho_p}} \;;\; \hat{\sigma}_Z^2 = \hat{\gamma}(0)\{1 - \underline{\widehat{\phi_p}}'\,\underline{\widehat{\rho_p}}\} = \hat{\gamma}(0)\{1 - \underline{\widehat{\rho_p}}'\hat{R}_p^{-1}\,\underline{\widehat{\rho_p}}\}$$

- **In practice, use recursive Durbin-Levinson Algorithm:**
- **The algorithm iteratively calculates** $\underline{\widehat{\phi_h}} = (\hat{\phi}_{h1}, \dots, \hat{\phi}_{hh})'$ **for h= 1, 2,…**
- **The last component** $\hat{\phi}_{hh} = \hat{\alpha}(h)$ **is PACF at lag h.**
- **If** $|\hat{\alpha}(h)| < 1.96/\sqrt{n}$ **for all h > p &** $|\hat{\alpha}(h)| \geq 1.96/\sqrt{n}$ **for h=p, choose**

  AR(p) model $X_t = \hat{\phi}_{p1} X_{t-1} + \hat{\phi}_{p2} X_{t-2} + \dots + \hat{\phi}_{pp} X_{t-p} + Z_t$
- **95% Confidence Interval for true parameter** ϕ_{pj}: $\hat{\phi}_{pj} \pm 1.96/\sqrt{n}\,\hat{v}_{jj}^{-1/2}$
- ***Iterative formula for*** $\hat{v}_{jj}$ ***is on the next slide.***

11.1.1. Yule-Walker Estimation for AR(p) models: The Durbin-Levinson Algorithm.

- **Yule-Walker estimates are solutions of these equations with sample acfs substituted for unknown acfs:**

$$\underline{\widehat{\phi}}_p = \widehat{R}_p^{-1}\underline{\widehat{\rho}}_p \;;\; \widehat{\sigma}_Z^2 = \widehat{\gamma}(0)\{1 - \underline{\widehat{\phi}}_p'\underline{\widehat{\rho}}_p\} = \widehat{\gamma}(0)\{1 - \underline{\widehat{\rho}}_p'\widehat{R}_p^{-1}\underline{\widehat{\rho}}_p\}$$

- The Durbin-Levinson algorithm formulas with following modifications–

– written through autocovariances $\gamma(= \gamma(0)\rho)$ rather than autocorrelations ρ,
– using sample estimates,
– using new notation (recursive) for the estimate of the noise variance σ_Z^2:

$$\hat{v}_h = \hat{\gamma}(0)\{1 - \underline{\hat{\phi}}_h'\,\underline{\hat{\rho}}_h\} \equiv \hat{\gamma}(0) - \underline{\hat{\phi}}_h'\underline{\hat{\gamma}}_h \text{ (the same as } \hat{v}_h = \hat{\gamma}(0)\{1 - \underline{\hat{\rho}}_h'\hat{R}_h^{-1}\,\underline{\hat{\rho}}_h\})$$

are as follows:

$\hat{\phi}_{hh} = [\hat{\gamma}(h) - \sum_{j=1}^{h-1}\hat{\phi}_{h-1,j}\hat{\gamma}(h-j)]\hat{v}_{h-1}^{-1}$;
$\hat{\phi}_{h,j} = \hat{\phi}_{h-1,j} - \hat{\phi}_{hh}\hat{\phi}_{h-1,h-j}$, $j = 1, \ldots, h-1$, and
$\hat{v}_h = \hat{v}_{h-1}[1 - \hat{\phi}_{hh}^2]$.

No need to memorize these formulas!

Note: $\hat{\phi}_{11} = \hat{\rho}(1) = \hat{\gamma}(1)/\hat{\gamma}(0)$ and $\hat{v}_0 = \hat{\gamma}(0)$.

Application of the Durbin-Levinson Algorithm for Forecasting for AR(p).

- **Observed data : $(X_1, X_2, \ldots, X_n) = (x_1, x_2, \ldots, x_n)$ comes from AR(p)**

  $$X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p} = Z_t, \quad Z_t \sim WN(0, \sigma_Z^2)$$

- **Goal: Forecast X_{n+1} (unknown).**

➢ **Write** $X_{n+1} = \phi_1 X_n + \ldots + \phi_p X_{n+1-p} + Z_{n+1}$

PLAN:

➢ **Take conditional expectation, given $X_1, X_2, \ldots, X_n$:**

$$\hat{X}_{n+1} = E\{X_{n+1} | X_1, X_2, \ldots, X_n\} = E\{\phi_1 X_n + \ldots + \phi_p X_{n+1-p} + Z_{n+1} | X_1, X_2, \ldots, X_n\}$$
$$= \phi_1 X_n + \ldots + \phi_p X_{n+1-p} \quad (**)$$

➢ **Mean-Square Error (m.s.e) of prediction:**

$$v_n = E(X_{n+1} - \hat{X}_{n+1})^2 = E(Z_{n+1})^2 = \sigma_Z^2.$$

- **Conclude: The Durbin-Levinson algorithm 11.1.1 provides a way for recursive estimation of both coefficients $\hat{\phi}_j$ in the forecast equation (**) and the forecast error $\hat{v}_n = \hat{\sigma}_Z^2$.**

Main points to take from Lecture 8:

- **Lectures 1 – 7 were mainly devoted to building models and investigating their second order properties.**
- **Lecture 8 starts discussing data analysis with introducing formulas for calculation of Sample Mean, Sample ACF, Sample PACF and their distributions.**
- **Although there is no need to memorize formulas, students are expected to be able to quickly find relevant material and be able to use it.**
- **Students must be able to write and interpret confidence intervals for acf & pacf; in particular, the Bartlett's formula.**
- **Identification of models based on comparing sample ACF/PACF with theoretical ACF/PACF derived for different models in lectures 1-7.**
- **AR(p): Yule-Walker estimation of model parameters;**
  The Durbin-Levinson Algorithm;
  Forecasting using the Durbin-Levinson Algorithm.

Some simple R commands used to create previous slides

- **To plot theoretical acf and pacf for pure SAR: X_t-0.8X_{t-12}=Z_t (seasonal with s=12, Φ_1=0.8):**

```
phi=c(rep(0,11), 0.8)                                   #pure seasonal AR(1) with phi=0.8
ACF=ARMAacf(ar=phi,ma = 0, 70) [-1]                             #[-1]to remove lag 0
PACF=ARMAacf(ar=phi,ma = 0, 50, pacf=TRUE)
plot(ACF, type="h", xlab="lag", ylim=c(-.4, .8), main="acf for ARMA(0,0)x(1,0)_12"); abline(h=0)
plot(PACF, type="h", xlab="lag", ylim=c(-.4, .8), main="pacf for ARMA(0,0)x(1,0)_12"); abline(h=0)
```

- **To simulate data from $SAR(1)_{12}$: X_t-0.8X_{t-12}= Z_t (seasonal with s=12, Φ_1=0.8):**

```
set.seed(90210)
phi=c(rep(0,11), 0.8)
ar120 <- arima.sim(list(ar=phi), n = 120, sd = 1)
```

- **To plot 70 theoretical ACF/PACF for SARMA(0,0,1)x(0,0,1):**

  $Y_t=(1+.7B)(1+.6B^{12})Z_t = Z_t + .7Z_{t-1} + .6Z_{t-12} + .42\, Z_{t-13}$, that is, MA(13) with few coefficients:

```
theacf=ARMAacf (ma =c(.7,0,0,0,0,0,0,0,0,0,0,.6,.42), lag.max=70)
plot(theacf, type="h", xlab="lag", ylim=c(-.4, .8), main="acf for ARMA(0,1)x(0,1)_12"); abline(h=0)
thepacf=ARMAacf (ma = c(.7,0,0,0,0,0,0,0,0,0,0,.6,.42),lag.max=70, pacf=T)
plot(thepacf, type="h", xlab="lag", ylim=c(-.4, .8), main="pacf for ARMA(0,1)x(0,1)_12"); abline(h=0)
```