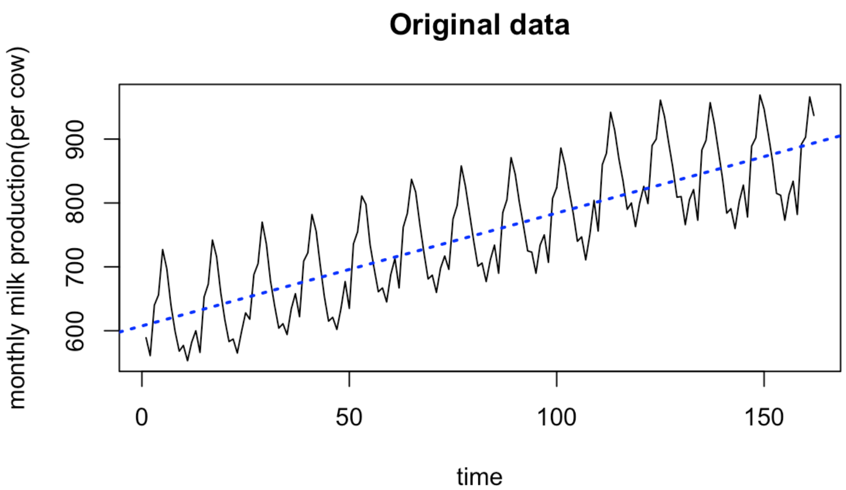
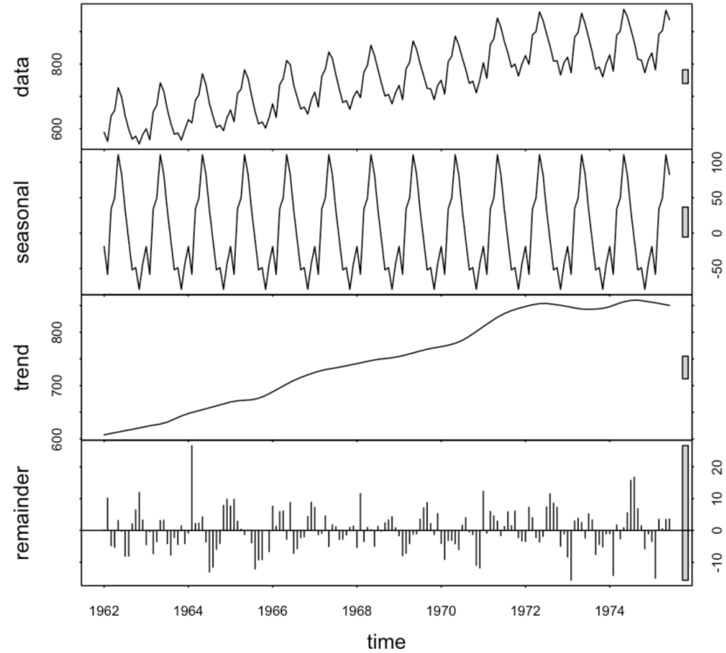
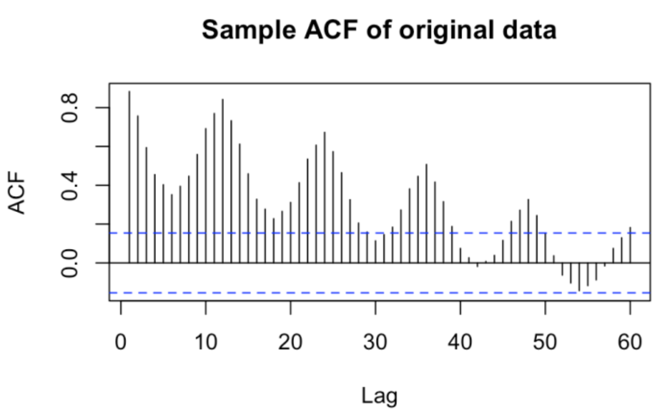
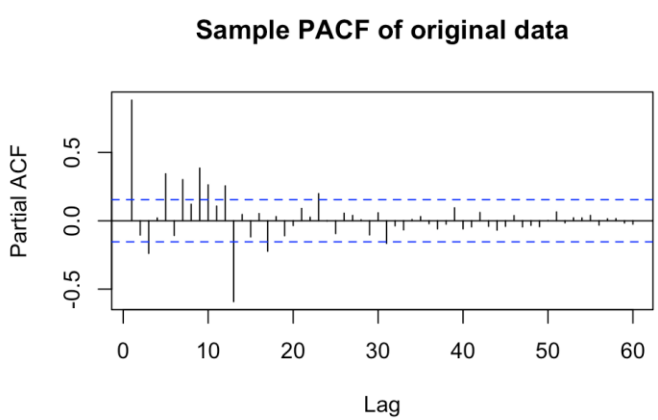
Examples for lecture 7:
nonstationarity seasonality differencing

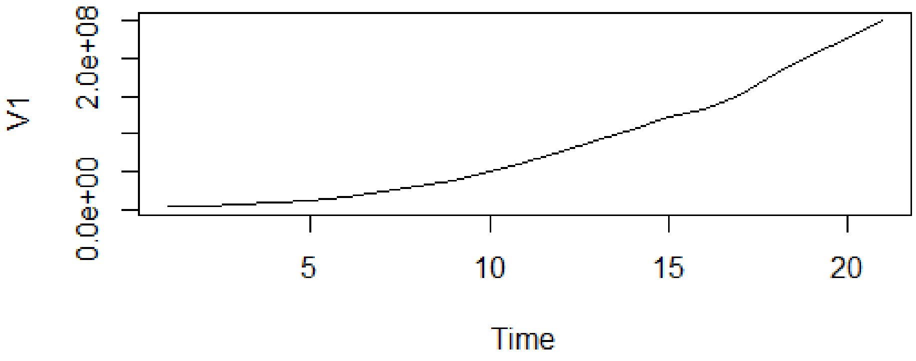






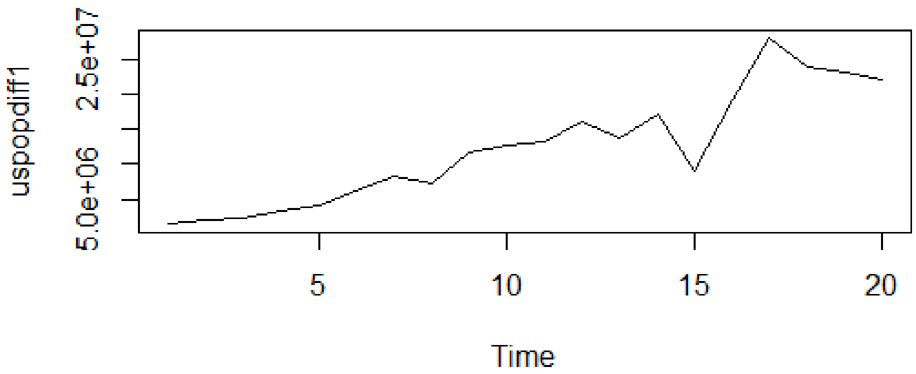


USPOP data file and its differences



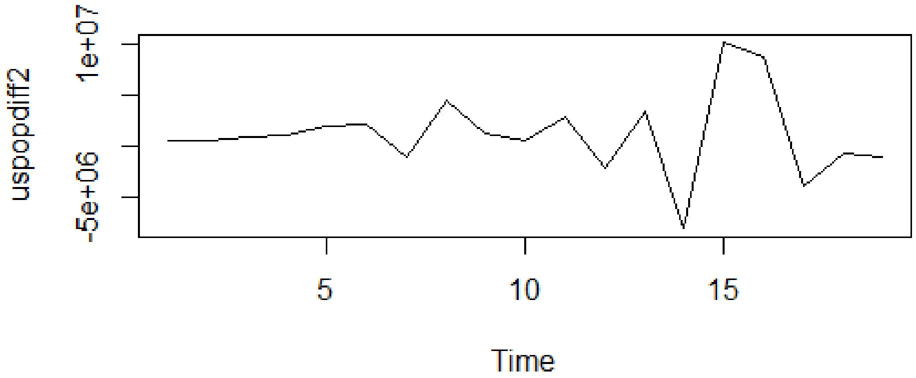
US Population, 21 values,

> var(uspop)
[1] 6.168983e+15



First difference of USPOP

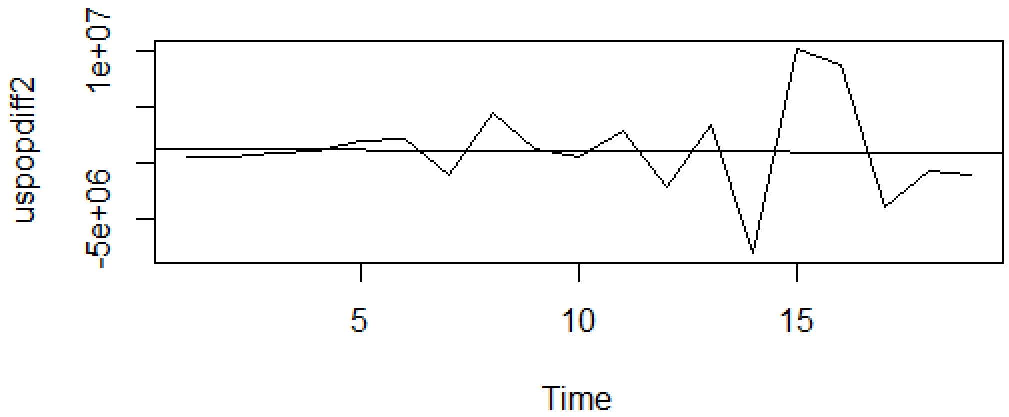
```
> uspopdiff1 <- diff(uspop,differences = 1);
> var(uspopdiff1)
[1] 6.597748e+13
```



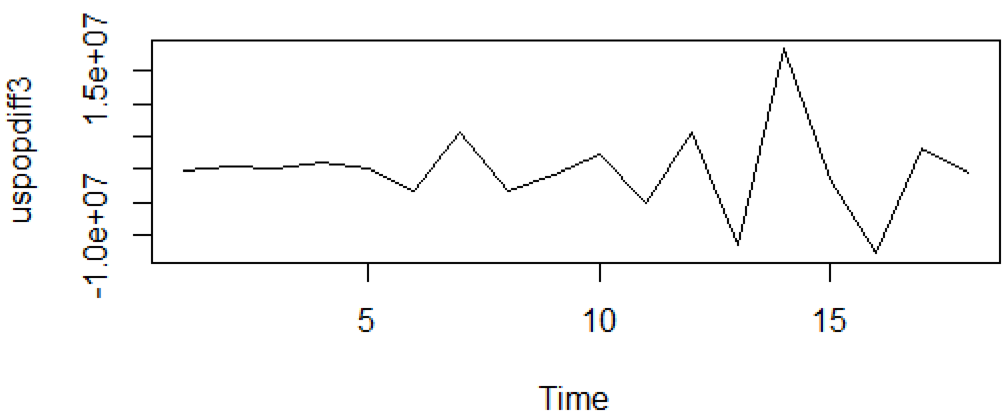
Second difference of USPOP;

```
>uspopdiff2 <- diff(uspopdiff1, differences = 1)
> plot.ts(uspopdiff2)
> var(uspopdiff2)
[1] 1.680728e+13
```

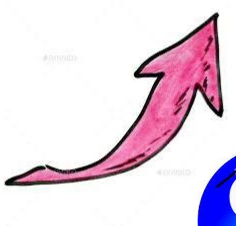
Second and third differences of USPOP



> var(uspopdiff2)
[1] 1.680728e+13

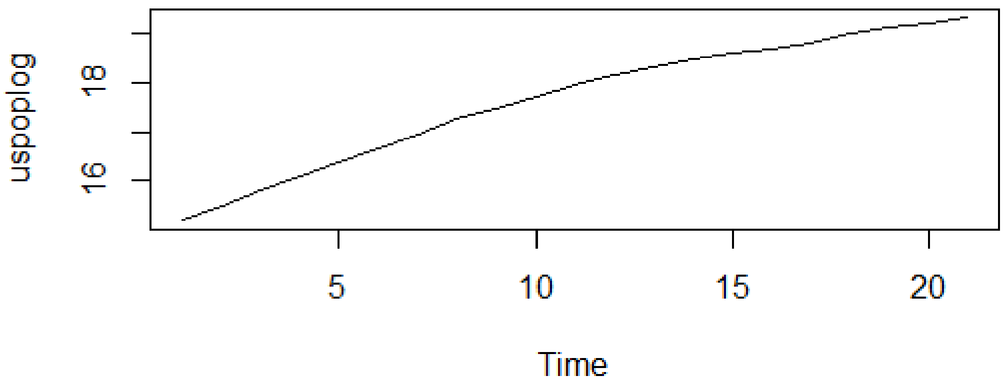


>var(uspopdiff3)
[1] 4.525136e+13



Note: Variance of the series decreased at differences 1 and 2 but increased when differencing the third time. What do you conclude?

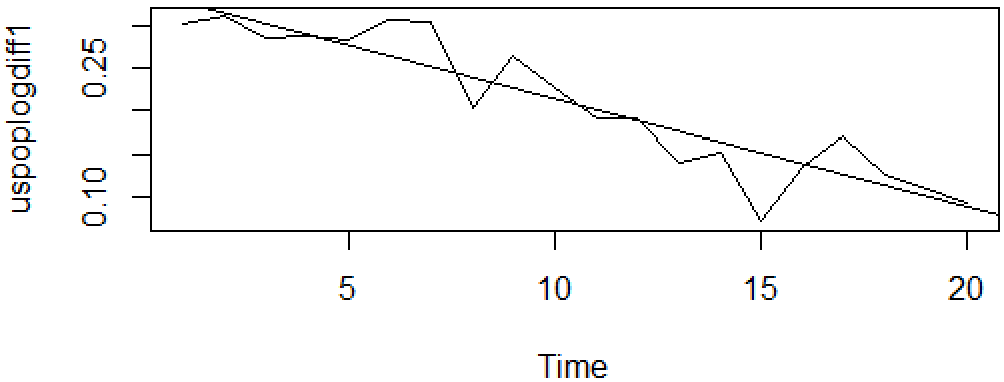
Log(USPOP) and its differences



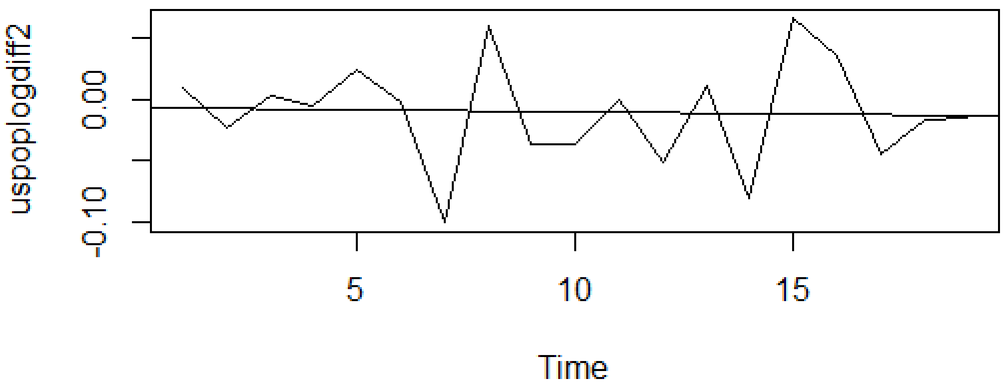
Plot of Log(USPOP)

> var(uspoplog)

[1] 1.716024



First difference of log(USPOP); sample variance is 0.006529038

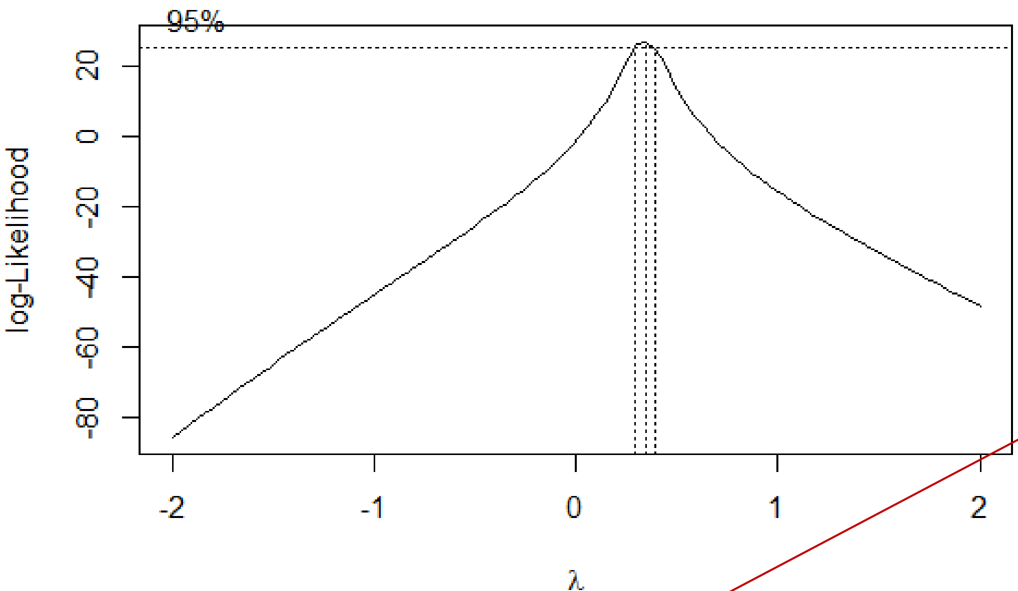


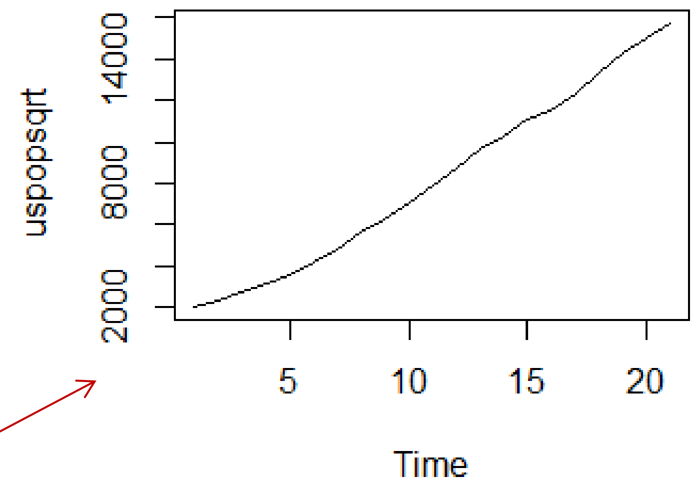
Second difference of log(USPOP); sample variance is 0.001799062

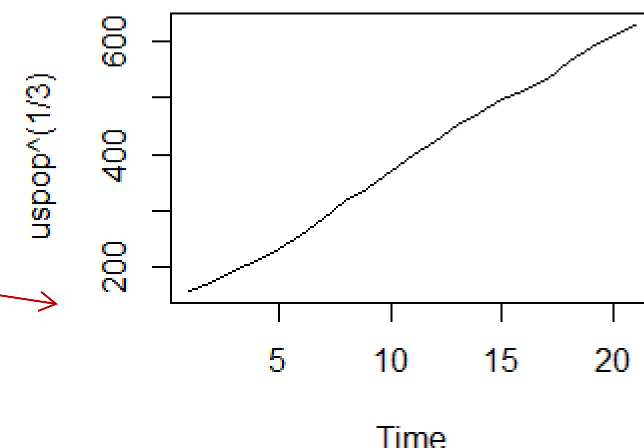
Box-Cox Transformation of USPOP

 # Instead of guessing, choose parameter λ of the Box-Cox transformation, use:

```
> require(MASS)
> bcTransform <- boxcox(uspop~ as.numeric(1:length(uspop)))          #plots the graph
> bcTransform$x[which(bcTransform$y == max(bcTransform$y))]       # gives the value of λ
  [1] 0.3434343
```

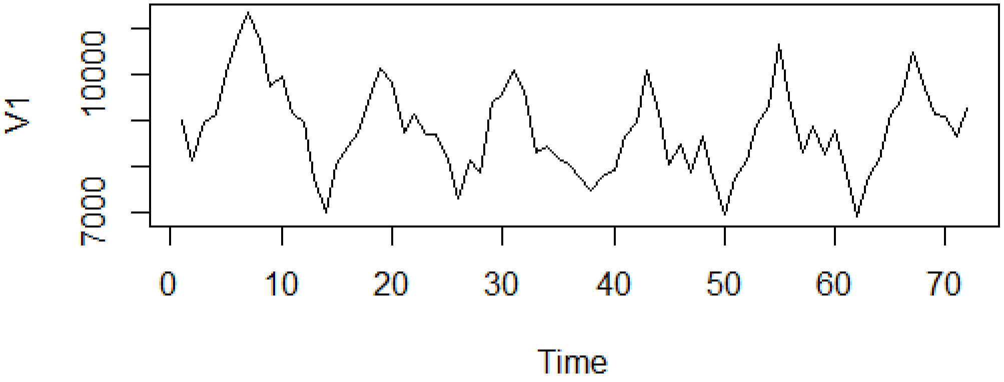




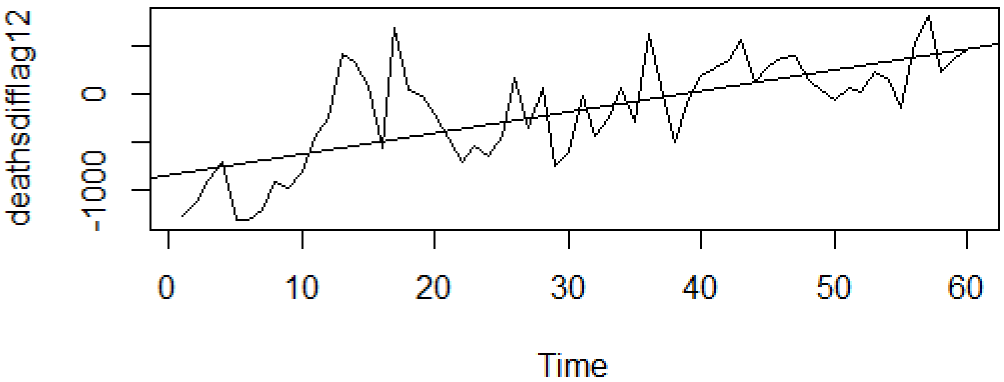


Try both λ=1/3 & λ=1/2 ,
i.e., square rand cubic root transformations
λ=1/2 could be in the confidence interval.
λ= 0 (log) not in the confidence interval.

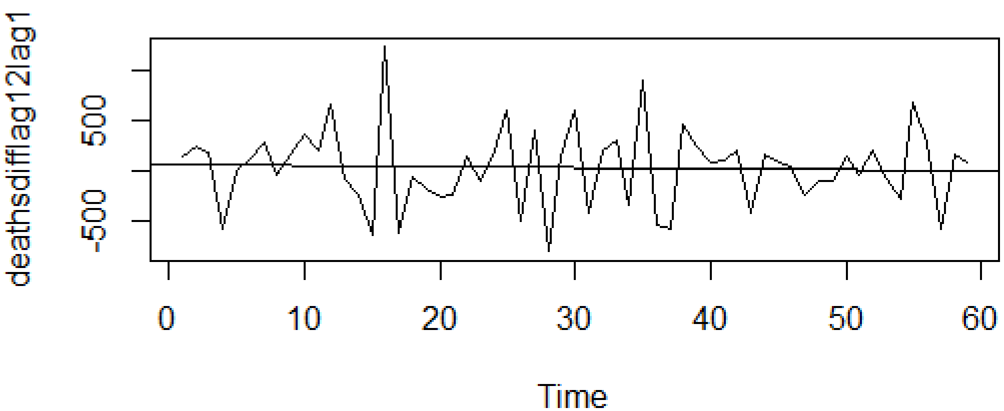
Accidental Deaths Data, differenced at lag 12 and then at lag 1



Original data: the monthly accidental deaths, 1973-1978
Sample variance: 918411.7

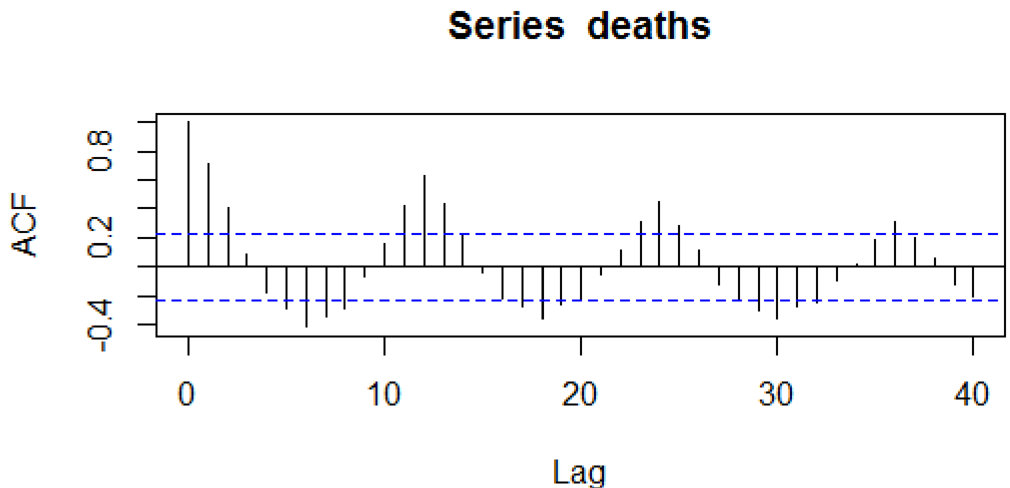


Difference at lag 12 to remove seasonality
Sample variance: 288714.5

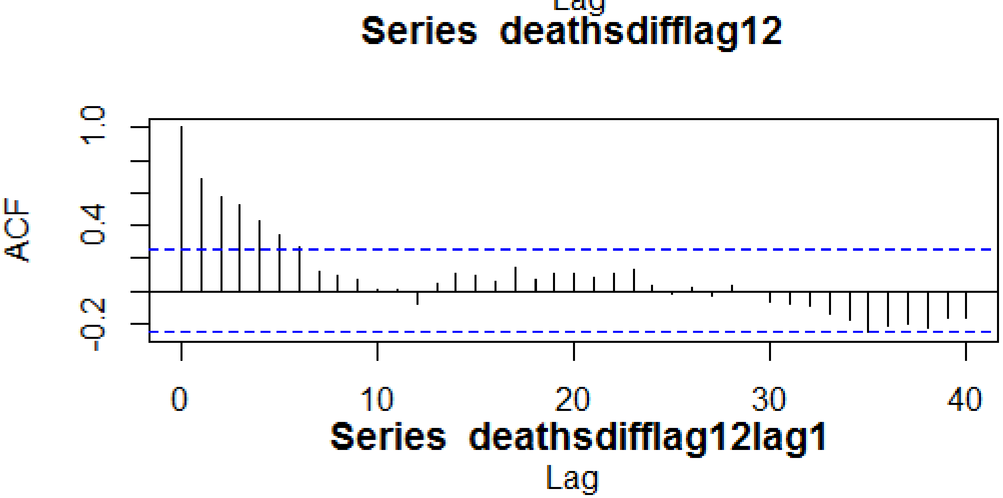


Differenced again at lag 1 to remove trend: $\nabla\nabla_{12} X_t$
Sample variance: 155301.9

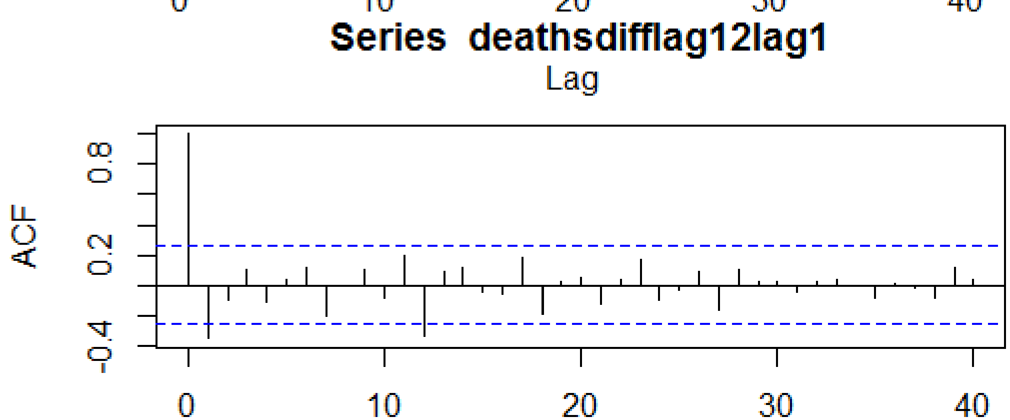
ACF for Accidental Deaths Data and its differences



ACF for original data: the monthly accidental deaths, 1973-1978



ACF for difference at lag 12



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

Seasonal ARIMA, s=12

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) :**

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

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)$_s$

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

Procedure to identify SARIMA:

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 $D = 1$.

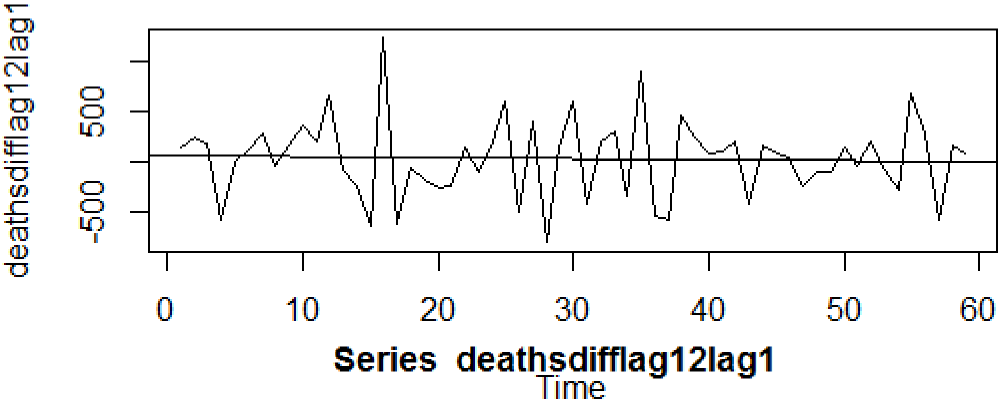
2. Find P and Q: look at $\hat{\rho}(ks)$, $k = 1, 2, \ldots$, i.e. look at ACF and PACF at lags which are multiples of s. Identify ARMA(P, Q).

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

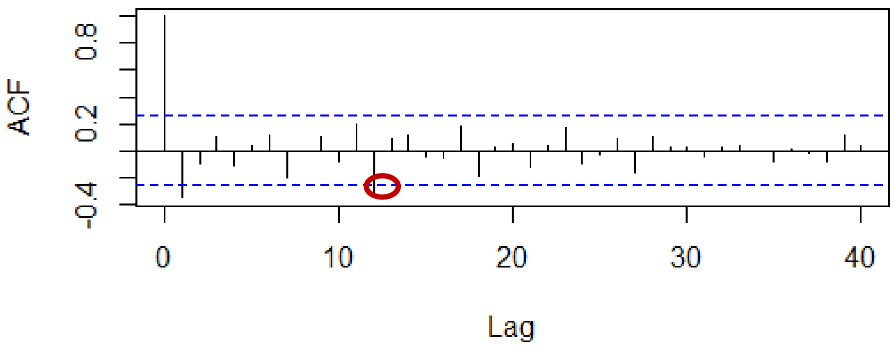
Note: Y_t constitutes ARMA$(p + sP, q + sQ)$ process in which some of the coefficients are zeros and the rest of the coefficients are functions of $\underline{\beta}' = (\underline{\phi}', \underline{\Phi}', \underline{\theta}', \underline{\Theta}')$.

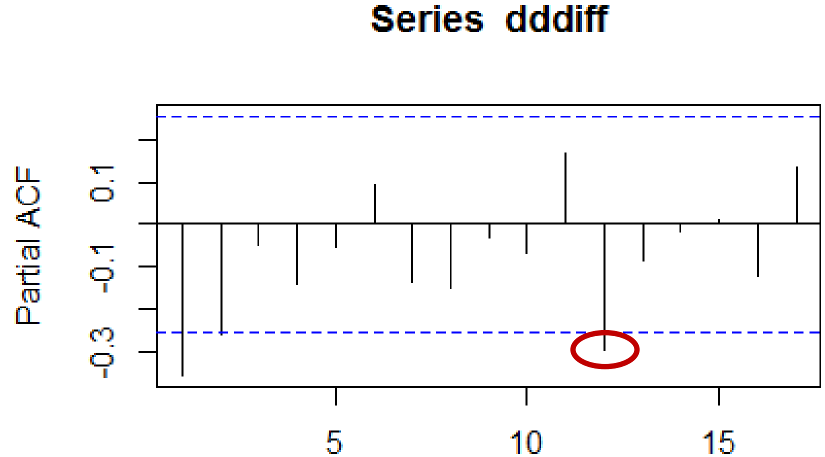
4. Use ML Estimation for $(\underline{\beta}, \sigma_Z^2)$ and use AICC and diagnostic checking to identify the best model.

ACF /PACF for Differenced Accidental Deaths Data



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





Lag

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

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

Observations: In SARIMA D = 1, d = 1

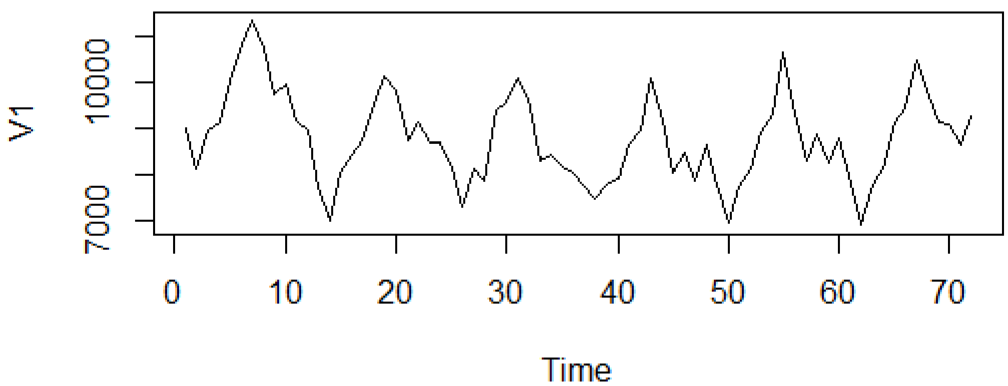
P/ACF large at lag 12; $\Rightarrow$ think SARIMA P=1 or Q=1 or P=Q=1

ACF $\rho(1) \neq 0, \rho(k)$, k=2, …, 12, within confidence intervals $\Rightarrow$ suspect MA, q=1

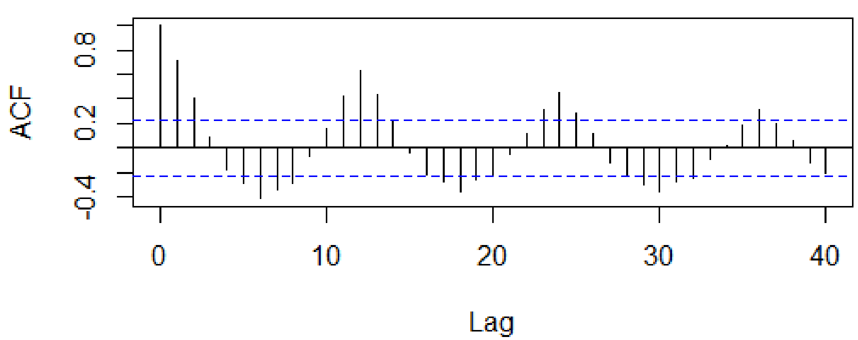
PACF $\alpha(1) \neq 0$, α (k) , k=2, …, 12, within confidence intervals $\Rightarrow$ suspect AR, p=1

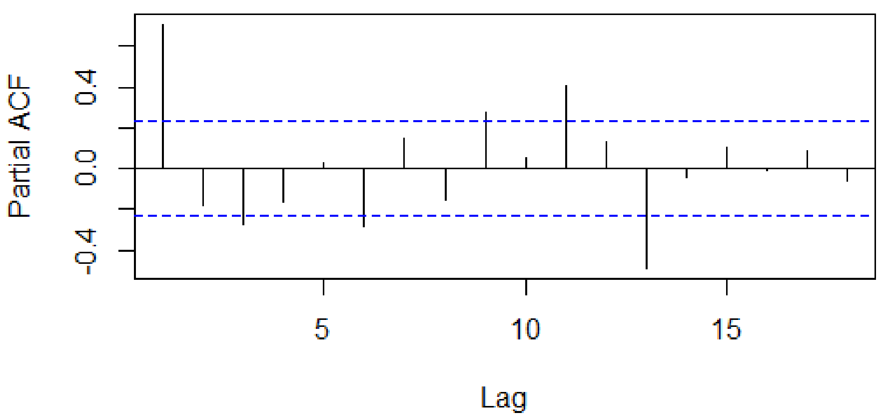
Also consider p=q=1.

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

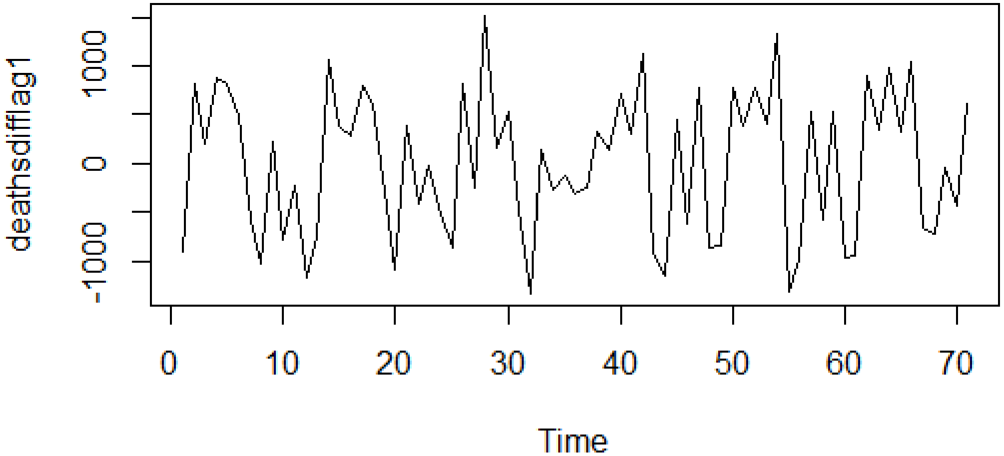


Seasonality not removed; acf periodic, remains large for large lags

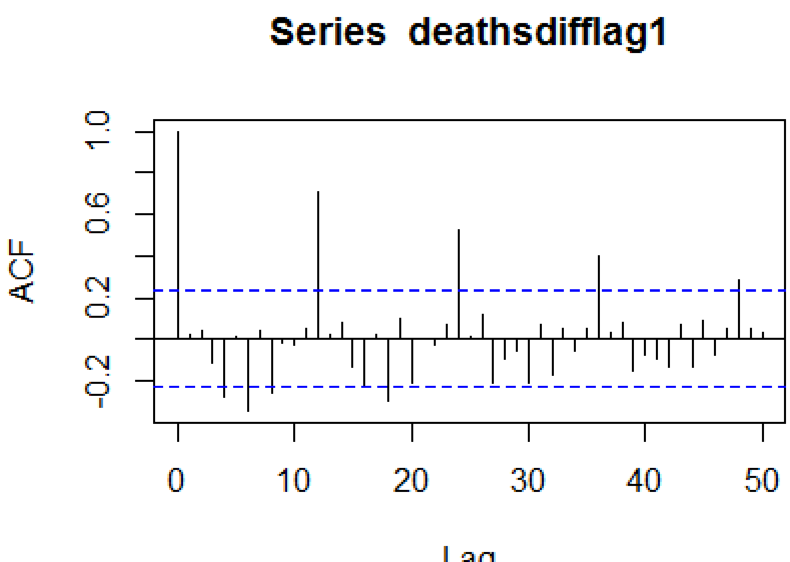


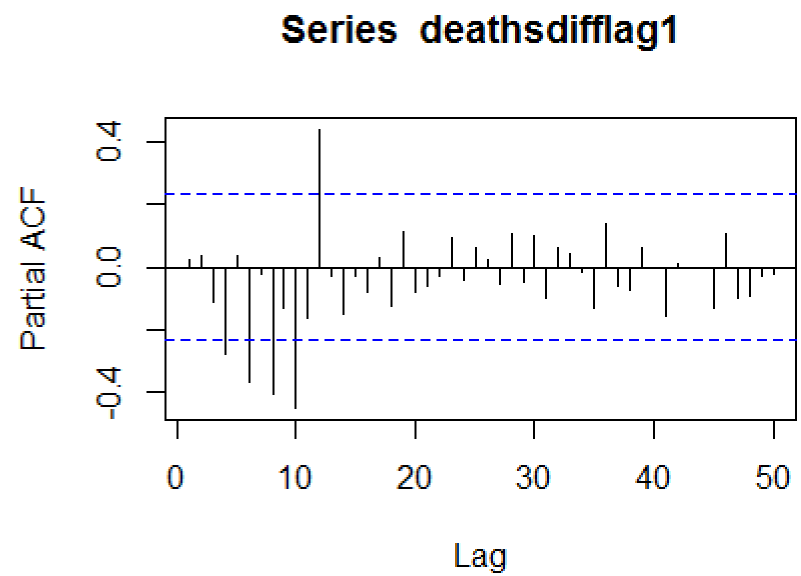


ACF for Accidental Deaths Data, Differenced at lag 1



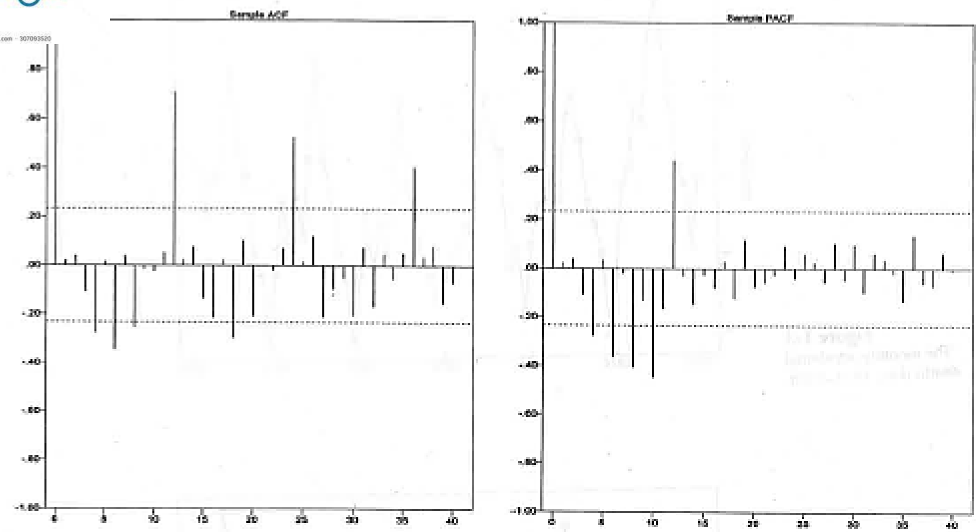
Seasonality not removed; acf periodic.



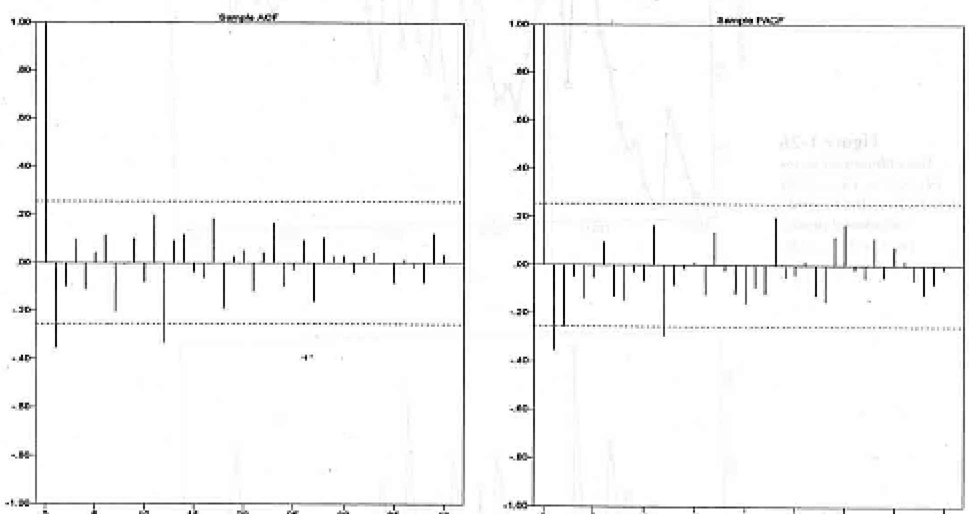


Compare with slide 8:
much harder to choose a model;
PACF suggests SAR at D=12, P=1, but monthly dependence p,q hard to determine

ACF for Accidental Deaths Data and its differences



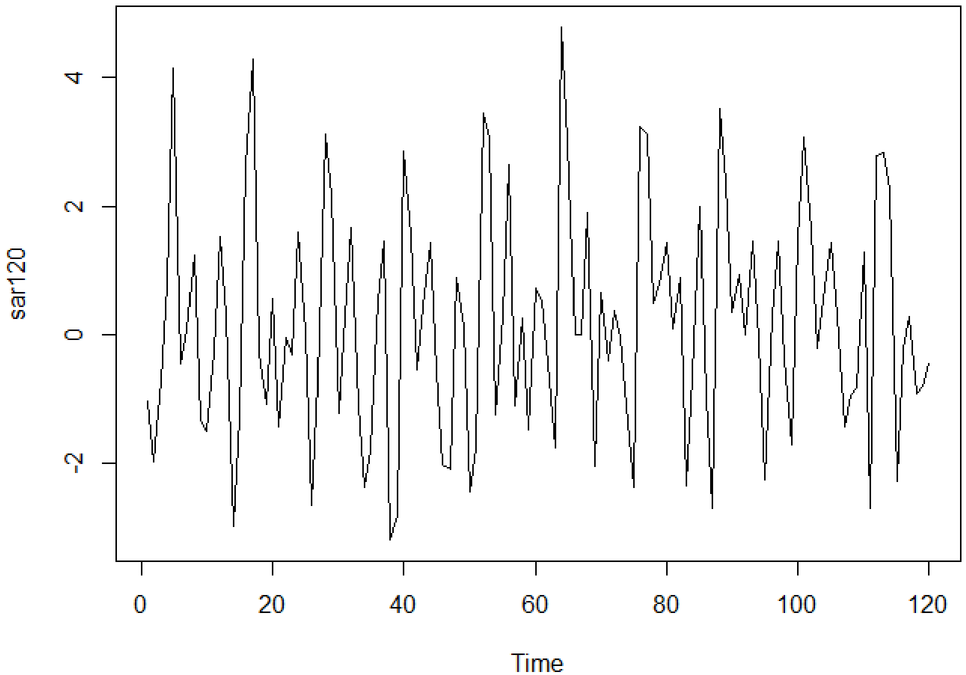
$\nabla(\text{Deaths})$



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

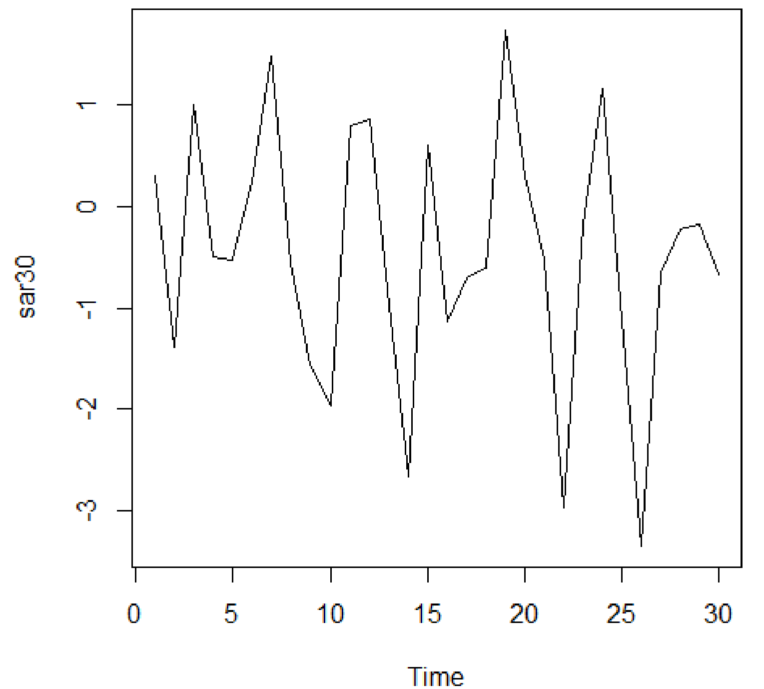
Note the difference:
Row 1: Seasonality not removed;
acf periodic and large
Row 2: Seasonality removed;
acf die out

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



10 periods: n=12 x 10

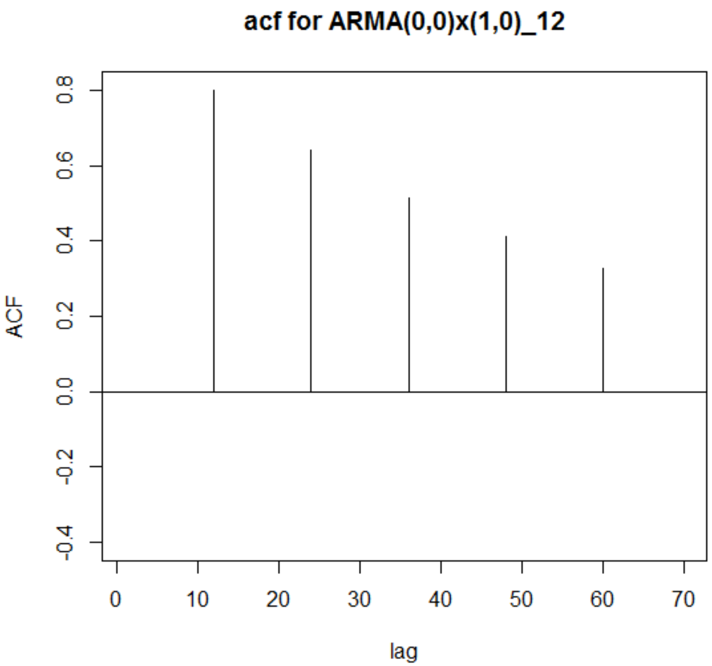
3 periods: n=12 x 3

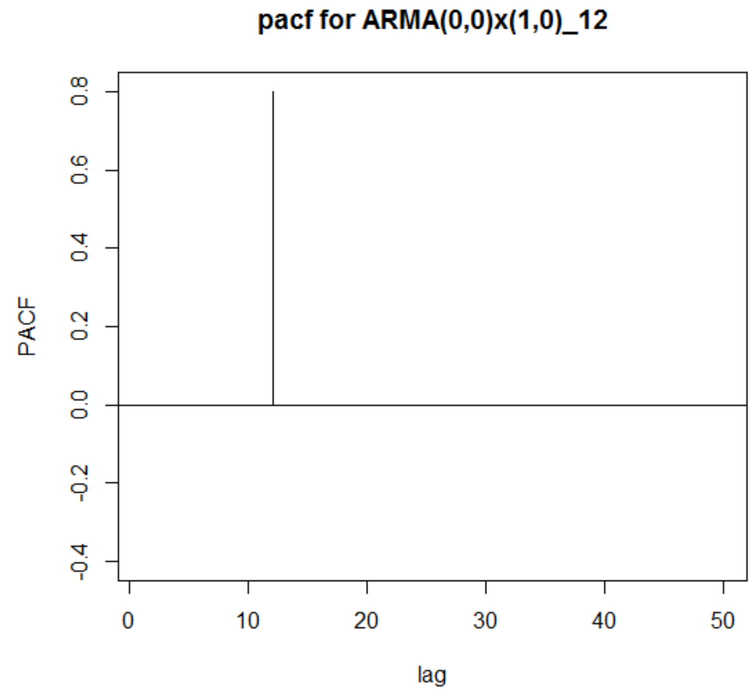


R code to simulate data:

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

Theoretical acf/pacf for SARIMA $(0,0)\times(1,0)_{12}$: $X_t - 0.8X_{t-12} = Z_t$.



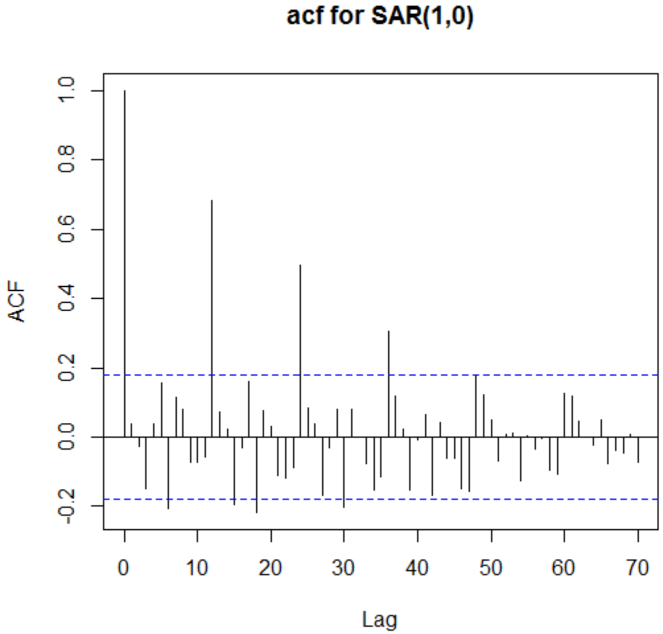


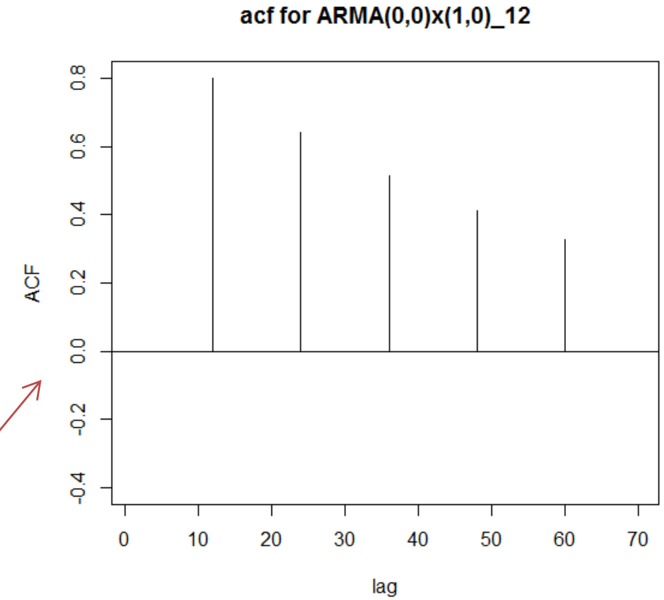
Annual AR: $\rho(12k) = \Phi_1^{\,k}$, k=1,2,…
Theoretical ACF looks like exponentially decaying spikes at lags 12, 24, 36, etc.

R code to plot theoretical acf and pacf for pure SAR: $X_t - 0.8X_{t-12} = Z_t$ (seasonal with s=12, $\Phi_1 = 0.8$):

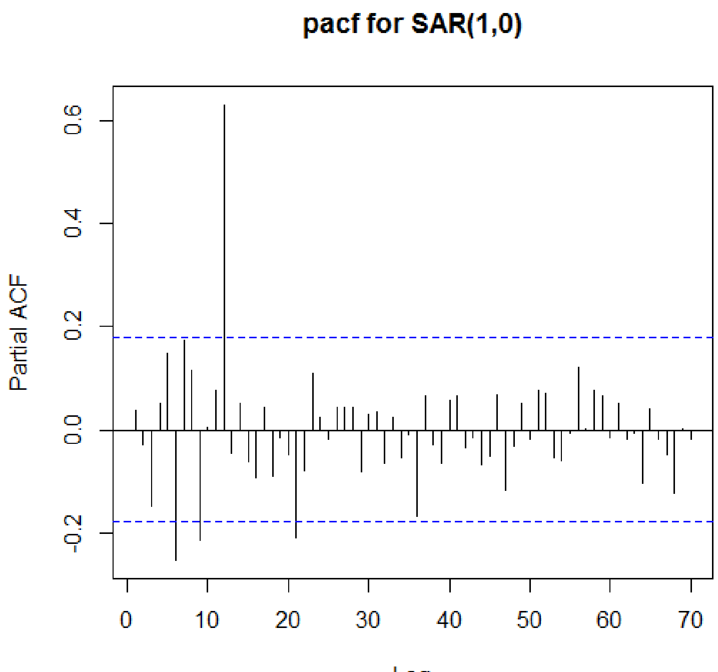
```
> phi=c(rep(0,11), 0.8)                    #pure seasonal AR(1) with phi=0.8
> ACF=ARMAacf(ar=phi,ma = 0, 70) [-1]      #[-1]to remove lag 0
> PACF=ARMAacf(ar=phi,ma = 0, 50, pacf=TRUE)
> plot(ACF, type="h", xlab="lag", ylim=c(-.4, .8), main="acf for ARMA(0,0)x(1,0)_12"); abline(h=0)
> plot(PACF,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

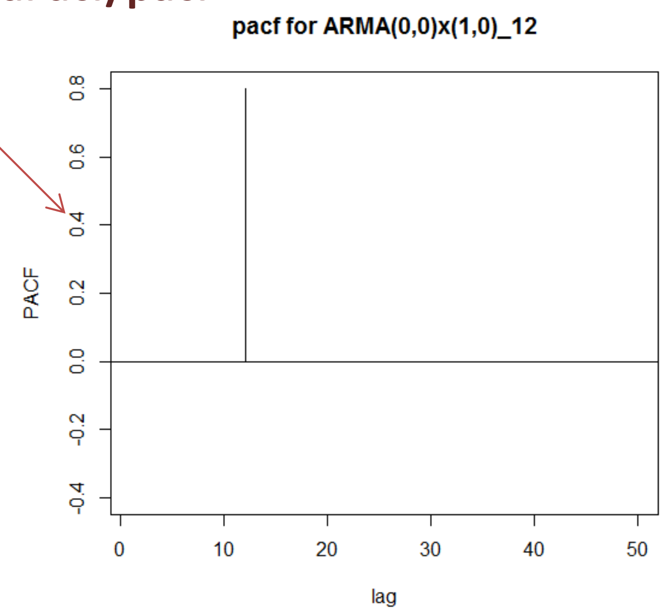
Sample acf/pacf for the simulated model: $X_t - 0.8X_{t-12} = Z_t$





theoretical acf/pacf





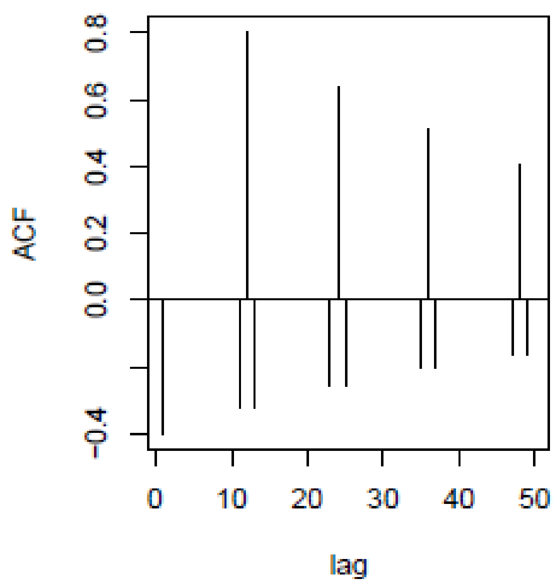
Behavior of the ACF and PACF for Pure SARMA Models

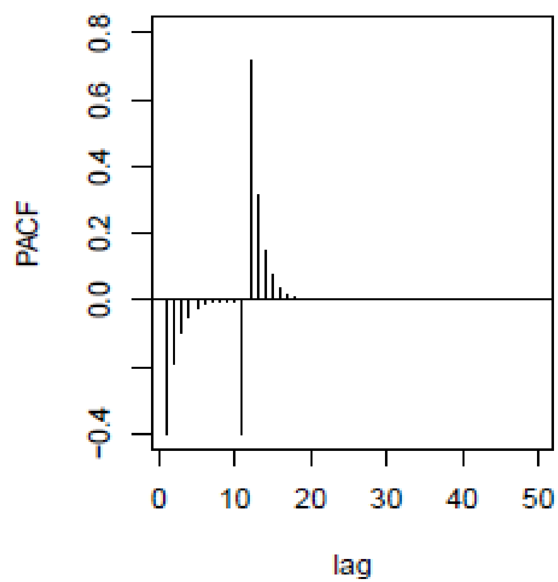
	$\text{AR}(P)_s$	$\text{MA}(Q)_s$	$\text{ARMA}(P, Q)_s$
ACF*	Tails off at lags ks, $k = 1, 2, \ldots,$	Cuts off after lag Qs	Tails off at lags ks
PACF*	Cuts off after lag Ps	Tails off at lags ks $k = 1, 2, \ldots,$	Tails off at lags ks

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

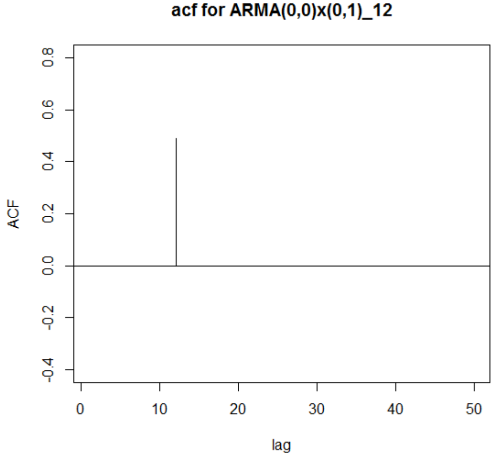
ACF and PACF of the mixed seasonal ARMA model $X_t - 0.8\, X_{t-12} = Z_t - 0.5\, Z_{t-1}$

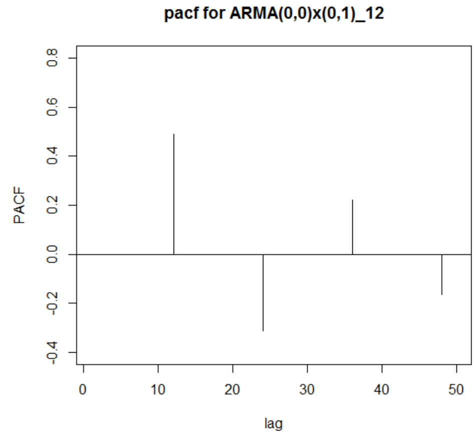
Here P=1, Q=0, p=0, q=1, s=12: $(1-0.8B^{12})X_t = (1 - 0.5B)\, Z_t$

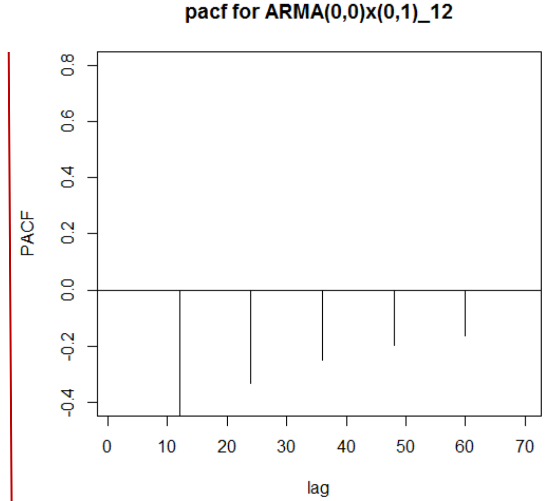


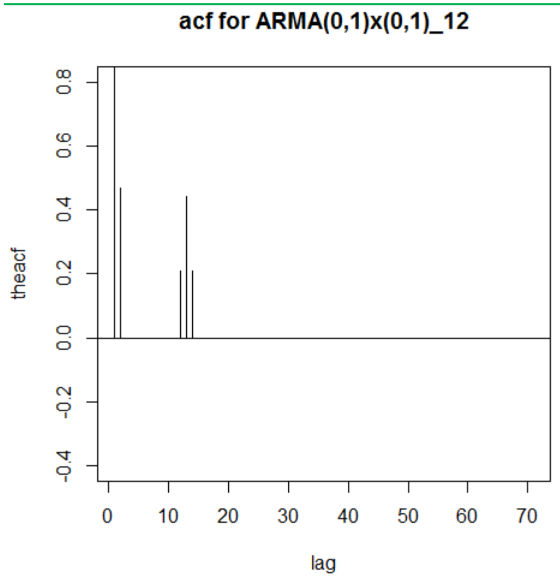


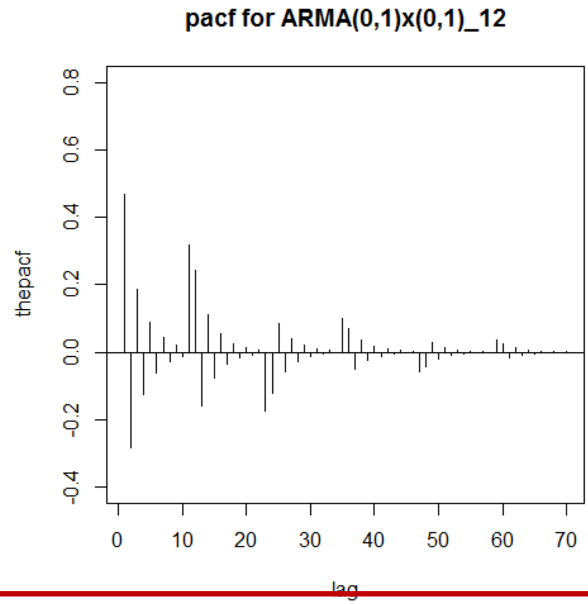
More ACF/PACF for SARIMA $(p,0,q)x(P,0,Q)_{12}$

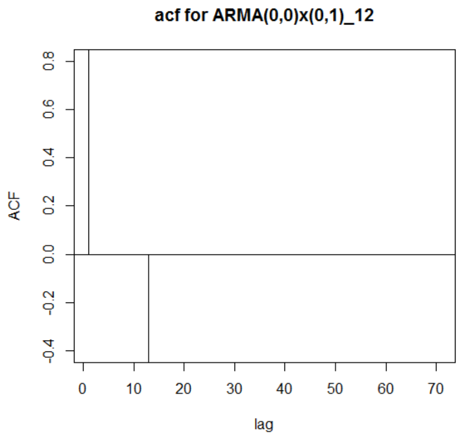








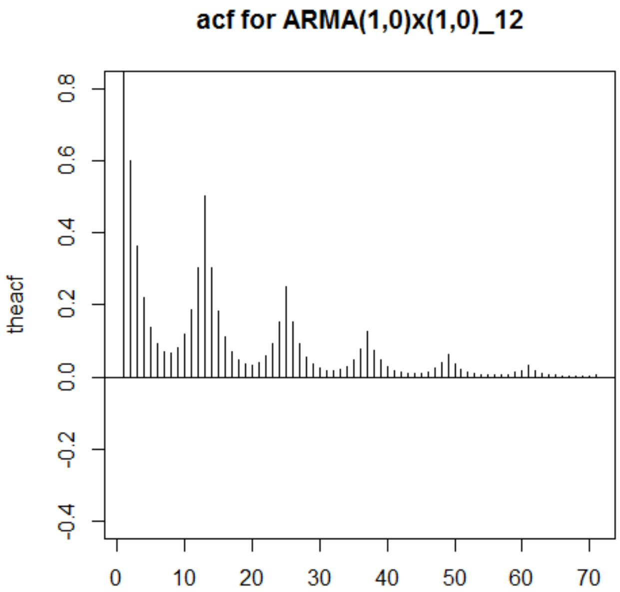


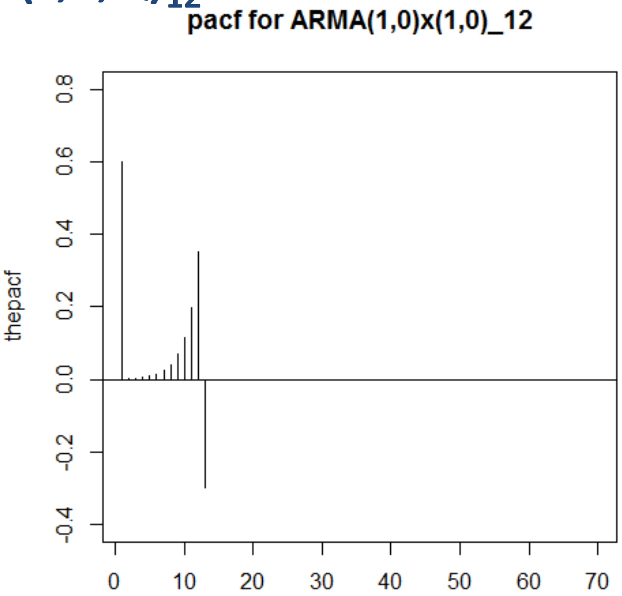


ARMA(0,0,1)x(0,0,1): $Y_t=(1+.7B)(1+.6B^{12})Z_t = Z_t + .6Z_{t-1} +.7\, Z_{t-12} + .42\, Z_{t-13}$

```
> 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)
```

More ACF/PACF for SARIMA $(p,0,q)\times(P,0,Q)_{12}$





<u>**Example**</u>: Let $Q = q = 0, P = p = 1, s = 12$. Model for $Y_t : (1-\phi_1 B)(1-\Phi_1 B^{12})Y_t = Z_t$, or

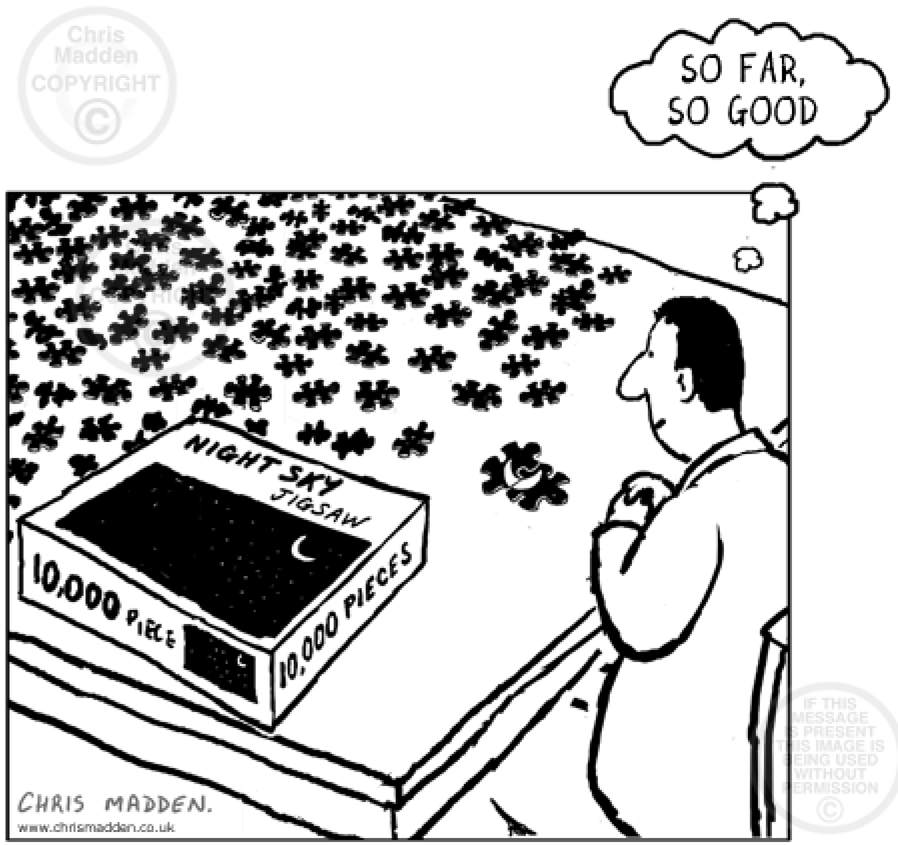
$$(1-\phi_1 B - \Phi_1 B^{12} + \phi_1\Phi_1 B^{13})Y_t = Z_t \; or \; Y_t - \phi_1 Y_{t-1} - \Phi Y_{t-12} - \phi_1\Phi_1 Y_{t-13} = Z_t$$

For $\phi = .6$ and $\Phi = .5$ we have: $Y_t - .6Y_{t-1} - .5Y_{t-12} - (-.3)Y_{t-13} = Z_t$ that is AR(13).
PACF has distinct spikes at lags 1, 12, 13 with a bit of action coming before lag 12. Then, it cuts off after lag 13.

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)
```

The End



SO FAR, SO GOOD
NIGHT SKY JIGSAW
10,000 PIECE
10,000 PIECES
CHRIS MADDEN.
www.chrismadden.co.uk

Good bye!