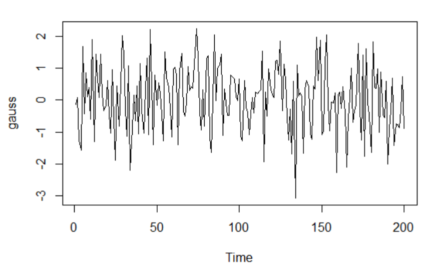
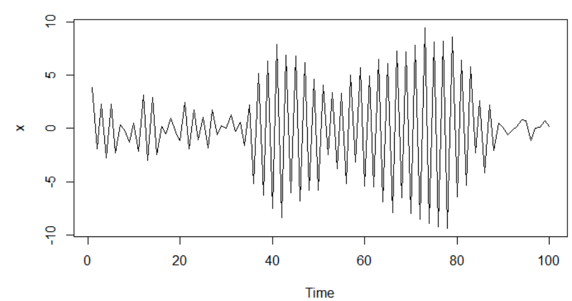
PSTAT 174/274 Week 3: Autoregressive and ARMA Models
PACF. Non-stationarity. Differencing. Transformations. Seasonality. ARIMA Models

Creating TS Models from White Noise and Past Observations





White Noise Z_t, t=1,2, ...

$X_t = \phi_1 X_{t-1} + Z_t, |\phi_1| < 1.$

Outline of Week 3 Lectures

Lecture 5: Autoregressive and ARMA Models

Lecture 6: PACF. Non-stationarity. Differencing. Transformations. Seasonality. ARIMA Models

REVIEW: Autoregressive models: AR(1)

$X_t = \phi_1 X_{t-1} + Z_t$ where $Z_t \sim WN(0, \sigma_Z^2)$ and $|\phi_1| < 1$.

- AR(1) is Always invertible: $Z_t = X_t - \phi_1 X_{t-1}$.
- AR(1) can be written as $X_t - \phi_1 X_{t-1} = Z_t$ that is $\boxed{(1 - \phi_1 B)X_t = Z_t}$
- When $|\phi_1| < 1$,

✓ Polynomial $\phi(z) = 1 - \phi_1 z$ has a root $z^* = 1/\phi_1$, $|z^*| > 1$

✓ for all $|z| \leq 1$, $|\phi_1| < 1$, $|\phi_1 z| < 1$, so that $\frac{1}{1-\phi_1 z} = 1 + \phi_1 z + \ldots + \phi_1^k z^k + \ldots$

✓ AR(1) has MA(∞) representation $X_t = Z_t + \phi_1 Z_{t-1} + \phi_1^2 Z_{t-2} + \cdots$

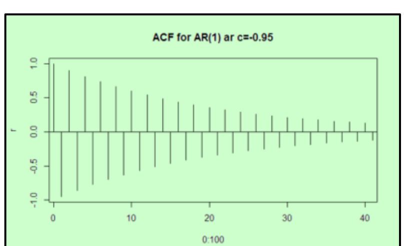
✓ AR(1) is stationary & causal *(b/c all MA processes are)*

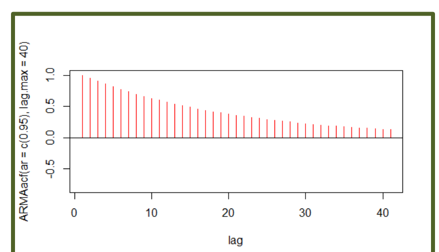
$B^k Z_s = Z_{s-k}$

✓ Variance of AR(1): $Var(X_t) = \gamma_X(0) = \frac{\sigma_Z^2}{1-\phi_1^2}$;

✓ ACF of AR(1): $\rho_X(1) = \phi_1,\ \rho_X(2) = \phi_1^2, \ldots, \rho_X(k) = \phi_1^k$.

Notice: ACFs of AR(1) decay exponentially, but are never zero!





>

REVIEW: 4.3 Autoregressive of Order p Models: AR(p)

$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + Z_t$ **where Z_t ~ WN $(0, \sigma_Z^2)$**

Properties of AR(p):

- **With ϕ *(z)* = $1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p$, the model can be rewritten as**

  ϕ ***(B)*** X_t = Z_t, Z_t ~ WN $(0, \sigma_Z^2)$
- ***AR(p) always invertible by its construction:*** $Z_t = \phi$ ***(B)*** X_t.
- ***AR(p) has MA(∞) representation, is stationary and causal when*** ϕ ***(z)*** $\neq 0$ **for** $|z| \leq 1$**, that is,**
- **the roots of the polynomial $\phi(z)$ lie outside of the unit circle.**

Next three slides discuss why the roots of the polynomial $\phi(z)$ have to be outside of the unit circle for an AR(p) process to have a MA representation.

NEW TODAY: ACF is found by solving Yule-Walker equations.

Some Algebra REVIEW

- ***z^* is called a root of a polynomial ϕ if $\phi(z^*) = 0$.***
- ***Polynomial ϕ of order p has p roots, $z_1, \ldots, z_p$, some might be the same number. Then, ϕ can be rewritten as***

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = \left(1 - \frac{1}{z_1} z\right)\ldots\left(1 - \frac{1}{z_p} z\right).$$

Check:

$\phi(z) = 0$ iff $\left(1 - \frac{1}{z_k} z\right) = 0$ for some k. This happens iff $z = z_k$ = one of the roots of ϕ

Example: *$\phi(z) = 1 - 2.5z + 2z^2 - 0.5z^3$ has $z_1 = z_2 = 1$ and $z_3 = 2$ and can be written as*

$$\phi(z) = \left(1 - \frac{1}{2} z\right)(1 - z)^2.$$

A fact from calculus: $\sum_{n=0}^{\infty} q^n$ is convergent iff $|q| < 1$. Then, $\frac{1}{1-q} = \sum_{n=0}^{\infty} q^n$.

Proof: As $N \to \infty$, $\sum_{n=0}^{N} q^n = \frac{1-q^{N+1}}{1-q} \to \frac{1}{1-q}$ iff $q^{N+1} \to 0$.

$q^{N+1} \to 0$ iff $|q| < 1$.

If $|q| > 1 \Rightarrow |q|^{N+1} \to \infty$, that is, the series does not converge.
If $q = 1$ the limit is $0/0$.

Some Algebra REVIEW and MA Representation for AR(p)

- **Inverting polynomials:**
- ***Let polynomial ϕ of order p have roots, $z_1, \ldots, z_p$. Then, ϕ can be rewritten as***

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = \left(1 - \frac{1}{z_1} z\right)\left(1 - \frac{1}{z_2} z\right) \ldots \left(1 - \frac{1}{z_p} z\right).$$

$$\text{Then, } \phi^{-1}(z) \equiv \frac{1}{\phi(z)} = \left\{\frac{1}{1-(1/z_1)z}\right\}\left\{\frac{1}{1-(1/z_2)z}\right\} \cdots \left\{\frac{1}{1-(1/z_p)z}\right\}$$

Each of the fractions can be represented by a convergent geometric series

$$\frac{1}{1-\frac{1}{z_k}z} = \sum_{n=0}^{\infty}\left(\frac{1}{z_k}z\right)^n,\ k = 1, 2, \ldots, p.$$

if $\left|\frac{1}{z_k} z\right| < 1$. This is guaranteed for all $|z| \leq 1$, if $\left|\frac{1}{z_k}\right| < 1 \Leftrightarrow |z_k| > 1$.

Conclude: *$\phi^{-1}(z)$ has a convergent infinite series representation if its roots $z_1, \ldots, z_p$ satisfy conditions:* $|z_k| > 1$, k = 1, 2, …, p. The infinite series is a product of geometric series corresponding to terms *$\left(1 - \frac{1}{z_k} z\right)^{-1}$*.

If $\phi^{-1}(z) = \sum_{j=0}^{\infty} \psi_j z^j$, then $\phi^{-1}(B) = \sum_{j=0}^{\infty} \psi_j B^j$ and therefore

X: $\phi(B) X_t = Z_t$ has MA(∞) representation $X_t = \phi^{-1}(B) Z_t = \sum_{j=0}^{\infty} \psi_j B^j Z_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$

Condition: All *roots $z_1, \ldots, z_p$ of polynomial ϕ satisfy conditions:* $|z_k| > 1$, k = 1, 2, …, p.

>

Some Algebra REVIEW

For complex roots, condition $|z_k|$ > 1, k = 1, …, p, means that the roots $z_1, \ldots, z_p$ of the polynomial $\phi(z)$ lie outside of the unit circle:

Math facts:

z* is a root of polynomial $\phi(z)$ if it satisfies the characteristic equation: ϕ (z*) = 0.

Roots of polynomial of order p > 1 may be complex and lie on a complex plane (x,y).

For z*= x+iy, condition |z*| = $|\sqrt{\{x^2 + y^2\}}|$ > 1 means that on the complex plane (x,y), z* lies is outside the unit circle $x^2 + y^2$ = 1. Points z: |z| < 1 are inside the circle.

Example:

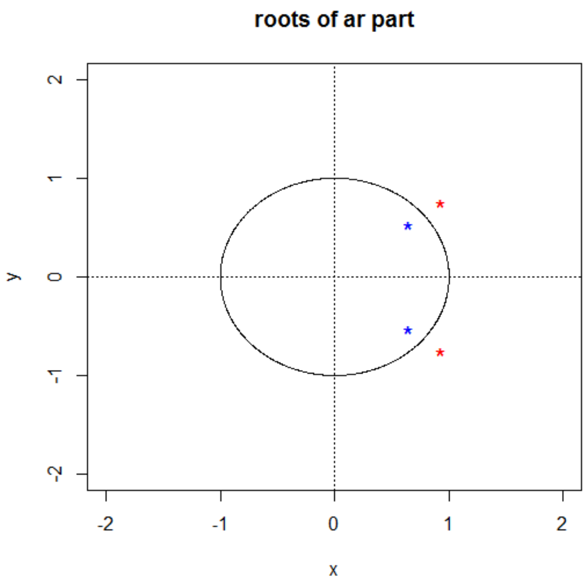
Roots of $\phi(z) = 1 - 1.3z + 0.7z^2$ *corresponding to* AR(2): $X_t = 1.3X_{t-1} - 0.7X_{t-2} + Z_t$.

Roots are complex and are outside unit circle:

```
> polyroot(c(1, -1.3, 0.7))
0.9285714+0.7525467i
0.9285714-0.7525467i
```

On the plot, these are red stars.

Blue stars are inversed roots $\frac{1}{z_k}$.



Check your understanding

A stationary autoregressive model of order one can be written as

$$Y_t = C + \phi_1 Y_{t-1} + Z_t$$

Determine which of the following statements about this model are false

(A) The parameter C must not equal 1.

(B) The absolute value of the parameter ϕ_1 must be less than 1.

(C) If the parameter $\phi_1 = 0$, then the model reduces to a white noise process.

(D) If the parameter $\phi_1 = 1$, then the model is a random walk.

(E) Mean of Y_t is C. (assume $\phi_1 \neq 0$, $|\phi_1| < 1$, $C \neq 0$).

check your answers on the next page

Check your understanding

A stationary autoregressive model of order one can be written as

$$Y_t = C + \phi_1 Y_{t-1} + Z_t$$

Determine which of the following statements about this model is false

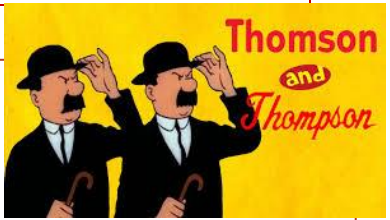
(A) The parameter C must not equal 1.

(B) The absolute value of the parameter ϕ_1 must be less than 1.

(C) If the parameter $\phi_1 = 0$, then the model reduces to a white noise process.

(D) If the parameter $\phi_1 = 1$, then the model is a random walk.

(E) Mean of Y_t is C. (assume $\phi_1 \neq 0$, $|\phi_1| < 1$, $C \neq 0$).



(A) is false: the intercept C can be any value.

(B) is correct: $|\phi_1| < 1$ is a condition for stationarity of AR(1).

(C) is correct: with $\phi_1 = 0$, $Y_t = C + Z_t$ is a white noise with mean C.

(D) is correct: $Y_t = C + Y_{t-1} + Z_t$ is a random walk with mean Ct:

$Y_1 = C + Z_1$; $Y_2 = C + (C + Z_1) + Z_2 = 2C + Z_1 + Z_2$; $Y_3 = C + (2C + Z_1 + Z_2) + Z_3 = 3C + Z_1 + Z_2 + Z_3$, etc

(E) is false: $E(Y_t) = C + \phi_1 E(Y_{t-1}) + 0$. B/c Y_t is stationary, $E(Y_{t-1}) = E(Y_t) = C/(1 - \phi_1)$

See you in part 2!

Outline of Lecture 5 Part II

New concepts:

- ***Yule-Walker equations to solve for ACVF and ACF of AR(p)***

AR(p): 4.3 ACF and Yule-Walker Equations

$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + Z_t$ **where** $Z_t \sim WN(0, \sigma_Z^2)$
or $\phi(B) X_t = Z_t$, $Z_t \sim WN(0, \sigma_Z^2)$, **with** $\phi(z) \neq 0$ **for** $|z| \leq 1$.

Task: Calculate ACVF (and ACF) of AR(p) Model:

$E(Z_t Z_s) = 0$, $t \neq s$.
$E(Z_t^2) = \sigma_Z^2$ for all t

$\gamma_X(k) = E(X_t X_{t-k})$

$= E\{(\phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + Z_t) X_{t-k}\}$

$= \phi_1 E(X_{t-1}X_{t-k}) + \phi_2 E(X_{t-2}X_{t-k}) + \ldots + \phi_p E(X_{t-p}X_{t-k}) + E(Z_t X_{t-k})$

- $E(X_{t-1}X_{t-k}) = \gamma_X(k-1)$; … ; $E(X_{t-p}X_{t-k}) = \gamma_X(k-p)$;

- $E(Z_t X_{t-k}) = 0$ if $k \geq 1$ **b/c X_{t-k} has MA representation and depends on Z_{t-k}, …:**
  $X_{t-k} = \sum_{j=0}^{\infty} \psi_j Z_{t-k-j} = Z_{t-k} + \psi_1 Z_{t-k-1} + \psi_2 Z_{t-k-2} + \ldots$

- $E(Z_t X_t) = \sigma_Z^2$ ($k=0$) **b/c X_t has MA representation and depends on Z_t, Z_{t-1}, …**
  (see calculations for AR(1) in Part II of Lecture 4 slides for Week 2 or § 4.1 of LNs)

$$\gamma_X(k) = \begin{cases} \phi_1 \gamma_X(k-1) + \phi_2 \gamma_X(k-2) + \ldots + \phi_p \gamma_X(k-p), & k \geq 1 \\ \phi_1 \gamma_X(1) + \phi_2 \gamma_X(2) + \ldots + \phi_p \gamma_X(p) + \sigma_Z^2, & k = 0 \end{cases}$$

AR(p): 4.3 ACF and Yule-Walker Equations

$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + Z_t$ **where** $Z_t \sim$ **WN** $(0, \sigma_Z^2)$
or $\phi(B) X_t = Z_t$, $Z_t \sim$ **WN** $(0, \sigma_Z^2)$, **with** $\phi(z) \neq 0$ **for** $|z| \leq 1$.

$$\gamma_X(k) = \begin{cases} \phi_1 \gamma_X(k-1) + \phi_2 \gamma_X(k-2) + \ldots + \phi_p \gamma_X(k-p), & k \geq 1 \\ \phi_1 \gamma_X(1) + \phi_2 \gamma_X(2) + \ldots + \phi_p \gamma_X(p) + \sigma_Z^2, & k = 0 \quad \text{(ACVF)} \end{cases}$$

ACF: $\rho_X(0) = 1$; $\rho_X(-k) = \rho_X(k) = \gamma_X(k)/\gamma_X(0)$.

$\rho_X(k) = \phi_1 \rho_X(k-1) + \phi_2 \rho_X(k-2) + \ldots + \phi_p \rho_X(k-p)$, $k \geq 1$.

Yule-Walker system of equations:

k=1: $\rho_X(1) = \phi_1 + \phi_2 \rho_X(1) + \ldots + \phi_p \rho_X(p-1)$;

k=2: $\rho_X(2) = \phi_1 \rho_X(1) + \phi_2 + \ldots + \phi_p \rho_X(p-2)$;

.....

k=p: $\rho_X(p) = \phi_1 \rho_X(p-1) + \phi_2 \rho_X(p-2) + \ldots + \phi_p$;

k=0 – next slide

AR(p): 4.3 ACF and Yule-Walker Equations

$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + Z_t$ where $Z_t \sim$ WN $(0, \sigma_Z^2)$
or $\phi(B) X_t = Z_t$, $Z_t \sim$ WN $(0, \sigma_Z^2)$, with $\phi(z) \neq 0$ for $|z| \leq 1$.

ACF: $\rho_X(0) = 1$; $\rho_X(-k) = \rho_X(k)$.

$\rho_X(k) = \phi_1 \rho_X(k-1) + \phi_2 \rho_X(k-2) + \ldots + \phi_p \rho_X(k-p)$, $k \geq 1$.

Yule-Walker equations for k = 0:

$$(**) \quad \gamma_X(0) = \phi_1 \gamma_X(1) + \phi_2 \gamma_X(2) + \ldots + \phi_p \gamma_X(p) + \sigma_Z^2$$

Recall: $\gamma_X(0) = \sigma_X^2$; $\rho_X(k) = \gamma_X(k)/\gamma_X(0)$.

So that $\gamma_X(k) = \rho_X(k)\,\gamma_X(0) = \rho_X(k)\,\sigma_X^2$.

Substitute into equation (**):

$$\sigma_X^2 = \phi_1 \rho_X(1)\,\sigma_X^2 + \phi_2 \rho_X(2)\,\sigma_X^2 + \ldots + \phi_p \rho_X(p)\,\sigma_X^2 + \sigma_Z^2$$

$$\sigma_X^2 = \frac{\sigma_Z^2}{1 - \phi_1 \rho_X(1) - \ldots - \phi_p \rho_X(p)}$$

4.3.3 Example: Yule-Walker Equations for AR(2)

$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t$ where $Z_t \sim$ WN $(0, \sigma_Z^2)$
or $\phi(B) X_t = Z_t$, $Z_t \sim$ WN $(0, \sigma_Z^2)$, with $\phi(z) \neq 0$ for $|z| \leq 1$.

ACVF for AR(2):

$$\gamma_X(k) = \begin{cases} \phi_1 \gamma_X(k-1) + \phi_2 \gamma_X(k-2), & k \geq 1 \\ \phi_1 \gamma_X(1) + \phi_2 \gamma_X(2) + \sigma_Z^2, & k = 0 \quad \text{(this is Var } \sigma_X^2\text{)} \end{cases}$$

ACF for AR(2): $\rho_X(k) = \phi_1 \rho_X(k-1) + \phi_2 \rho_X(k-2)$, $k \geq 1$.

Yule-Walker system of equations for AR(2):

k=1: $\rho_X(1) = \phi_1 + \phi_2 \rho_X(1)$

k=2: $\rho_X(2) = \phi_1 \rho_X(1) + \phi_2$;

Thus, $\rho_X(1) = \dfrac{\phi_1}{1-\phi_2}$; $\quad \rho_X(2) = \dfrac{\phi_1^2}{1-\phi_2} + \phi_2 = \dfrac{\phi_1^2 + \phi_2(1-\phi_2)}{1-\phi_2}$

$$\sigma_X^2 = \frac{\sigma_Z^2}{1 - \phi_1 \rho_X(1) - \phi_2 \rho_X(2)}$$ --Variance and ACF for AR(2)

Check your Understanding: Yule-Walker Equations for AR(2)

For a second-order autoregressive process, you are given:

$\rho_X(1) = 0.53$ & $\rho_X(2) = -0.22$.

(i) Determine coefficients of the model ϕ_1 and ϕ_2 .

(ii) Write model equation in two different forms;

(iii) Find $\rho_X(3)$.

check your answers on the next page

A Random Thought:

The goal of the class is to fit a time series dataset into a good model and use it to forecast future data points.
The above problem seems to have no relation to this goal. Is it a busywork?

An Answer (Not so Random) – Not a busywork:

In practice, this is exactly what R software will do for you:
-- from the given dataset, an R program (e.g., R command *acf*) will calculate sample ACF;
-- using Yule-Walker equations, an R program (e.g., R command *ar*) will calculate estimates of the model coefficients ϕ_1 and ϕ_2 .

Check your Understanding: Yule-Walker Equations for AR(2)

For a second-order autoregressive process, you are given:

$\rho_X(1) = 0.53$ & $\rho_X(2) = -0.22$.

(i) Determine coefficients of the model ϕ_1 and ϕ_2.

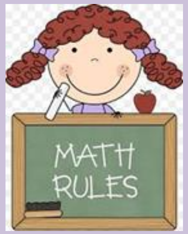
(ii) Write model equation in two different forms;

(iii) Find $\rho_X(3)$.

(i) Yule-Walker equations for AR(2) are (see example 4.3.3 of week 3)

$\rho_X(1) = \phi_1 + \phi_2\, \rho_X(1)$

$\rho_X(2) = \phi_1\, \rho_X(1) + \phi_2$.



Substitute $\rho_X(1) = 0.53$ and $\rho_X(2) = -0.22$, to get $0.53 = \phi_1 + \phi_2\,(0.53)$

$-0.22 = \phi_1\,(0.53) + \phi_2$.

Solve for ϕ_1 and ϕ_2:

-- from the second equation: $\phi_2 = -0.22 - \phi_1\,(0.53)$.

-- substitute to the first equation: $0.53 = \phi_1 + \{-0.22 - \phi_1\,(0.53)\}\,(0.53)$

$= \phi_1 + (-0.22)(0.53) - \phi_1\,(0.53)^2$.

-- calculate: $\phi_1 = \{(0.53)(1.22)\} / \{1 - (0.53)^2\} = 0.899$; $\phi_2 = -0.22 - (0.899)(0.53) = -0.696$.

(ii) AR(2) equation can be written as $X_t = 0.899\, X_{t-1} - 0.696\, X_{t-2} + Z_t$ **with $Z_t \sim$ WN $(0, \sigma_Z^2)$ or**

$X_t - 0.899\, X_{t-1} + 0.696\, X_{t-2} = Z_t$, **$Z_t \sim$ WN $(0, \sigma_Z^2)$ or**

$\phi(B)\, X_t = Z_t$, **$Z_t \sim$ WN $(0, \sigma_Z^2)$, where** $\phi(B) = 1 - 0.899\, B + 0.696\, B^2$.

(iii) From Yule-Walker equations for AR(2), $\rho_X(k) = \phi_1 \rho_X(k-1) + \phi_2\, \rho_X(k-2)$, **$k \geq 1$, we get:**

$\rho_X(3) = \phi_1 \rho_X(2) + \phi_2\, \rho_X(1) = (0.899)(-0.22) + (-0.696)(0.53) = -0.567$.

Check Your Understanding: MA(1)

You fit an invertible first-order moving average model to a time series.

The lag-one sample autocorrelation coefficient is -0.35.

Determine an initial guess for θ_1, the moving average parameter.

One way to think about this problem:

We have MA(1) equation: $X_t = Z_t + \theta_1 Z_{t-1}$; θ_1 unknown.

(i) From theory: For MA(1) model, ACF $\rho_X(1) = \theta_1/(1+\theta_1{}^2)$ depends on θ_1.

(ii) Given in this problem (in practice, estimated from data, see future lectures): $\rho_X(1) = -0.35$.

To get an answer, connect (i) and (ii) to find θ_1.

Check your answer of the next slide

Check Your Understanding: MA(1)

You fit an invertible first-order moving average model to a time series.

The lag-one sample autocorrelation coefficient is -0.35.

Determine an initial guess for θ_1, the moving average parameter.

We have MA(1) equation: $X_t = Z_t + \theta Z_{t-1}$; θ is unknown.

(i) From theory: For MA(1) model, ACF $\rho_X(1) = \theta/(1+\theta^2)$ depends on θ.

(ii) Given in this problem (in practice, estimated from data, see future lectures):

$\rho_X(1) = -0.35$.

To get an answer, connect (i) and (ii) to find θ :

$$\theta/(1+\theta^2) = -0.35 \text{ or } 0.35\,\theta^2 + \theta + 0.35 = 0.$$

This quadratic equation has two roots:

$$\theta = \{-1 + \sqrt{1 - 4(0.35)^2}\}/\{2(0.35)\} = -0.41,\ \theta = \{-1 - \sqrt{1 - 4(0.35)^2}\}/\{2(0.35)\} = -2.44$$

We are told that the time series is invertible, which for MA(1) means that $|\theta| < 1$.

Conclude: $\theta = -0.41$.

Quadratic formula to find roots of the quadratic equation $ax^2 + bx + c = 0$, is $x_{1,2} = \dfrac{-b \pm \sqrt{b^2-4ac}}{2a}$

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** $Y_t = 0.9\,Y_{t-1} + Z_t + 0.6\,Z_{t-1} - 0.3Z_{t-2}$

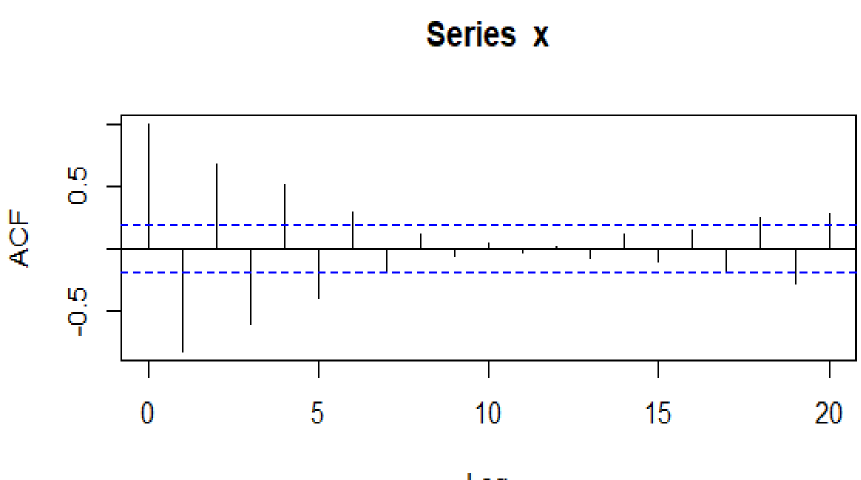
B. $X_t = -0.9\,X_{t-1} + Z_t$ **and** $Y_t = Z_t + 0.6\,Z_{t-1} - 0.3Z_{t-2}$

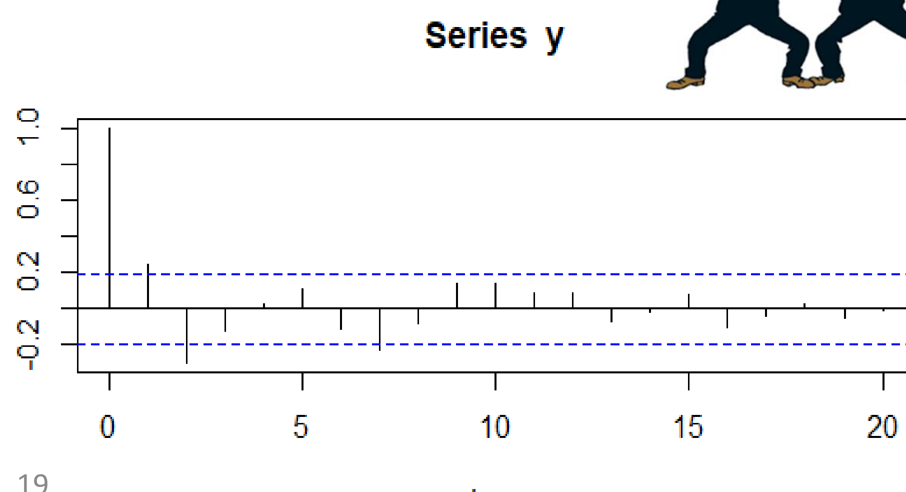
C. $X_t = 0.9\,X_{t-1} + Z_t$ **and** $Y_t = Z_t + 0.6\,Z_{t-1}$

D. $X_t = -0.9\,X_{t-1} + Z_t$ **and** $Y_t = Z_t + 0.6\,Z_{t-1} - 0.3Z_{t-2} + 0.4\,Z_{t-3}$

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

Choose one





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** $Y_t = 0.9\, Y_{t-1} + Z_t + 0.6\, Z_{t-1} - 0.3 Z_{t-2}$

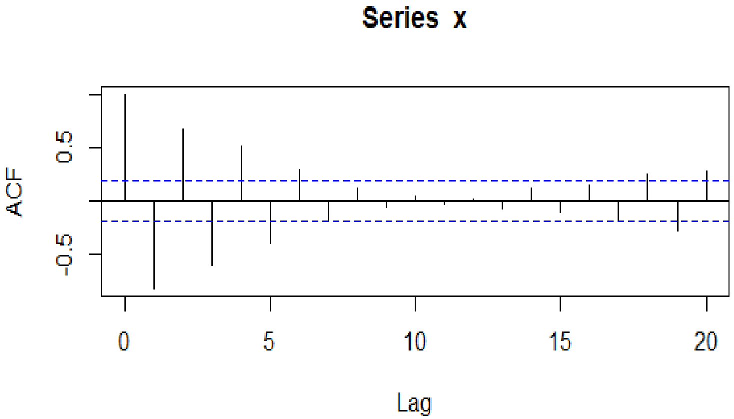
B. $X_t = -0.9\, X_{t-1} + Z_t$ **and** $Y_t = Z_t + 0.6\, Z_{t-1} - 0.3 Z_{t-2}$

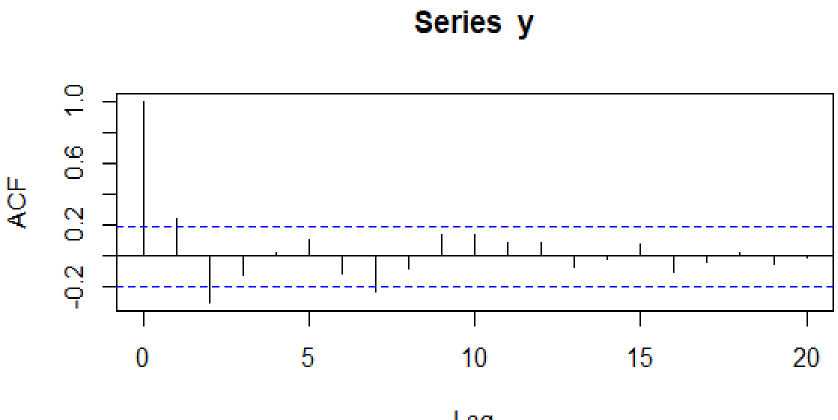
C. $X_t = 0.9\, X_{t-1} + Z_t$ **and** $Y_t = Z_t + 0.6\, Z_{t-1}$

D. $X_t = -0.9\, X_{t-1} + Z_t$ **and** $Y_t = Z_t + 0.6\, Z_{t-1} - 0.3 Z_{t-2} + 0.4\, Z_{t-3}$

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

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 ACF of Y, Y is MA(2), because $\hat{\rho}_Y(k) \approx 0$ for k >2. $\Rightarrow$ Answer: (B).

Main Points to Take from Part II of Lecture 5

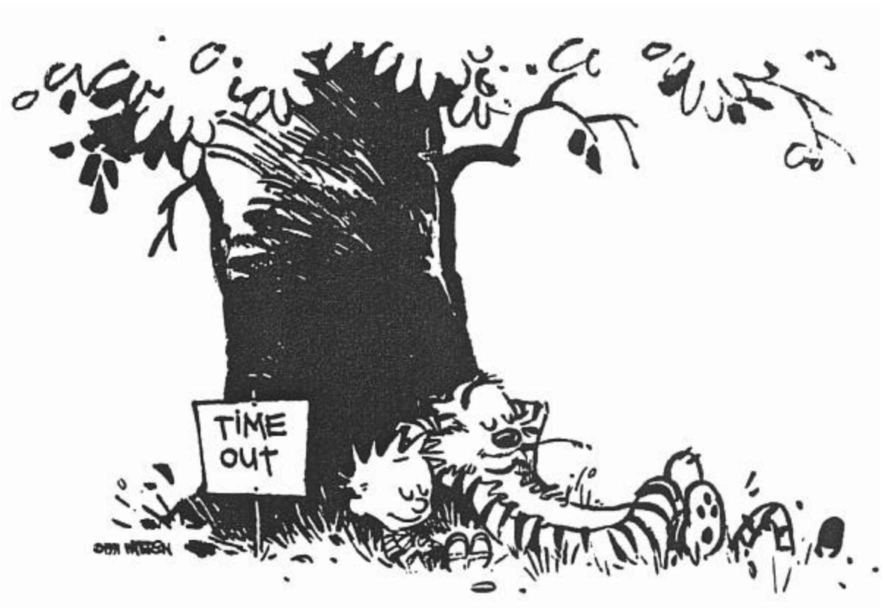
Autoregressive Models AR(p):

- **Model equation:** $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + Z_t$

  also $X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \ldots - \phi_p X_{t-p} = Z_t$

  also ϕ **(B)** $X_t = Z_t$
- **AR models are always invertible.**
- ***AR models are stationary when they have MA(∞) representation.***

  Restriction on coefficients: |z*| > 1 for z* such that ϕ (z*) = 0.
- **ACVF and ACF of AR(p) are obtained from the Yule-Walker equations.**
- **Example of Yule – Walker equations are their solution for AR(2).**



End of Part 2 of Lecture 5.

Take a Break and Move to Part 3!

Outline of Lecture 5 Part III

Mixed Autoregressive-Moving Average models ARMA

New concepts:

- ***Mixed Autoregressive-Aoving Average ARMA (p, q) models***

5.1 Mixed Autoregressive – Moving Average models: ARMA(1, 1)

$X_t - \phi_1 X_{t-1} = Z_t + \theta_1 Z_{t-1}$ where $Z_t \sim WN(0, \sigma_Z^2)$
$|\theta_1| < 1$ and $|\phi_1| < 1$.

Representation using polynomials and shift operator B: $B X_t = X_{t-1}$

With $\phi(z) = 1 - \phi_1 z$; $\theta(z) = 1 + \theta_1 z$,

Notice signs of θ 's and ϕ's

ARMA(1, 1): $\phi(B) X_t = \theta(B) Z_t$

ARMA(1, 1) is invertible if $|\theta_1| < 1$: $Z_t = \theta^{-1}(B)\, \phi(B)\, X_t$.
ARMA (1, 1) is stationary, causal, has MA(∞) representation if $|\phi_1| < 1$:
$X_t = \phi^{-1}(B)\, \theta(B)\, Z_t$

Coefficients of MA (∞) representation:

$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$ with $\psi_0 = 1$; $\psi_j = \phi_1^{j-1}(\phi_1 + \theta_1)$ for $j > 0$.

Derivation is in 5.1 section of lecture notes for week 3.

5.1 ACF of ARMA(1, 1)

$X_t - \phi_1 X_{t-1} = Z_t + \theta_1 Z_{t-1}$ where $Z_t \sim WN(0, \sigma_Z^2)$
$|\theta_1| < 1$ and $|\phi_1| < 1$.

ACF for ARMA(1, 1): $\rho_X(k) = \dfrac{(\phi_1 + \theta_1)(1 + \phi_1 \theta_1)}{(1 + 2\phi_1 \theta_1 + \theta_1^2)} \phi_1^{k-1}$, $k > 0$.

Derivation of ACVF using MA (∞) representation:

$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$ with $\psi_0 = 1$; $\psi_j = \phi_1^{j-1}(\phi_1 + \theta_1)$ for $j > 0$.

$\gamma_X(k) = E(X_t X_{t+k}) = E\{\sum_{j=0}^{\infty} \psi_j Z_{t-j} \cdot \sum_{m=0}^{\infty} \psi_m Z_{t+k-m}\}$

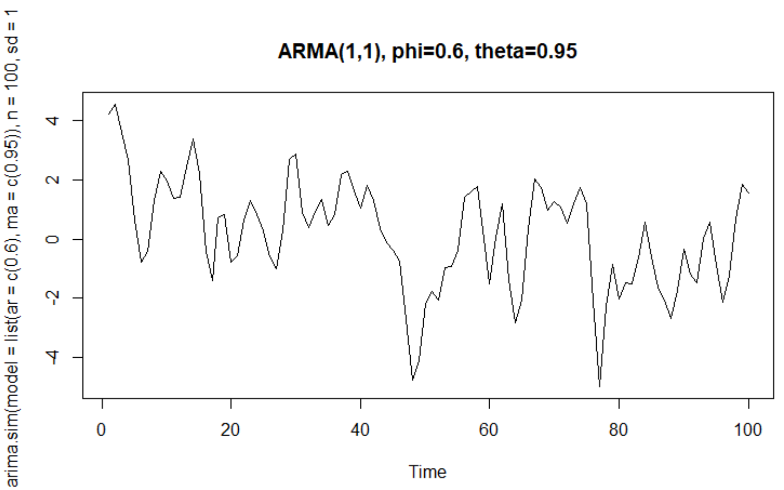
$= \sum_{j=0}^{\infty} \sum_{m=0}^{\infty} \psi_j \psi_m E\{Z_{t-j} Z_{t+k-m}\}$

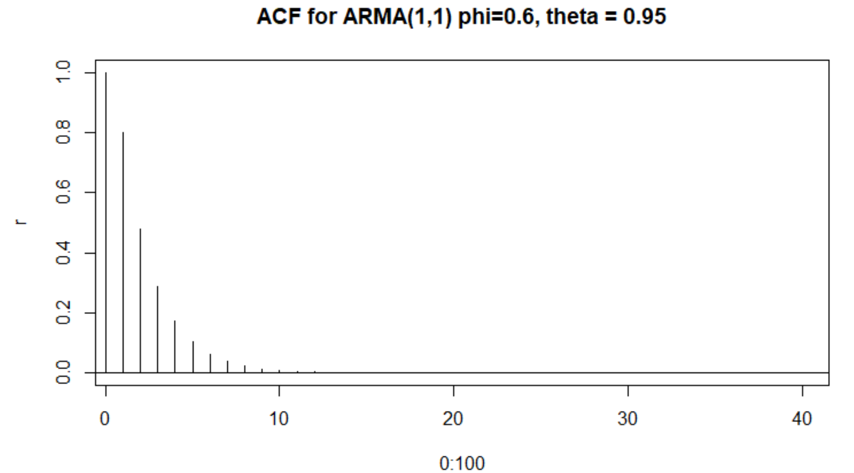
$= \sigma_Z^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k}$

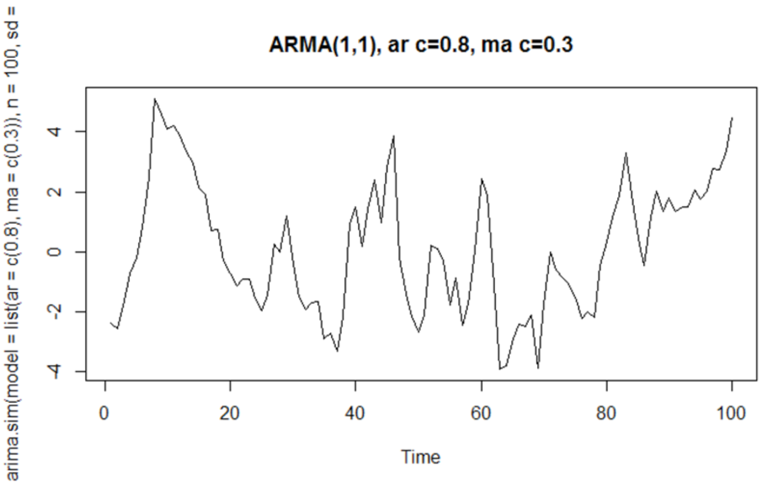
$E(Z_t Z_s) = 0$, $t \neq s$.
$E(Z_t^2) = \sigma_Z^2$ for all t

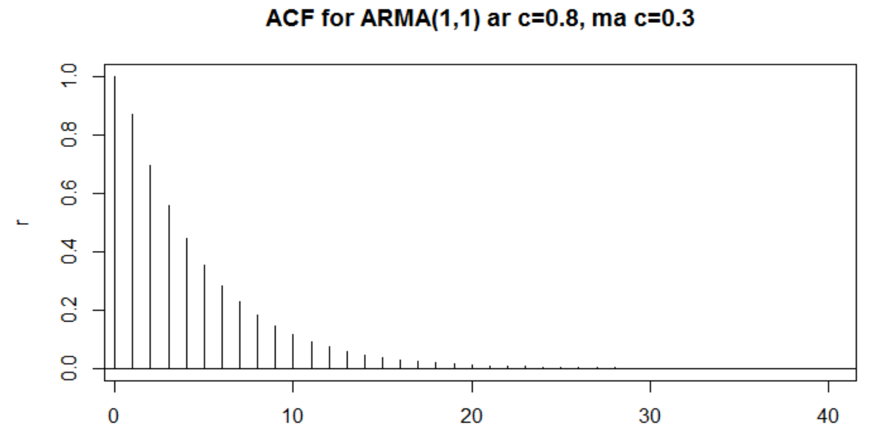
b/c $E(Z_{t-j} Z_{t+k-m}) = \sigma_Z^2$ iff $t - j = t + k - m$, that is, $m = k + j$.

Simulated values of ARMA(1,1) and corresponding ACF









Typical acf for ARMA(1,1): Spike at lag 1 (as in MA(1)), Exponential decay (as in AR(1)). Hard to see …

R commands used:

```
> plot.ts(arima.sim(model=list(ar=c(0.6), ma=c(0.95)), n = 100, sd = 1), main="ARMA(1,1), phi=0.6, theta=0.95")
> plot(0:100,ARMAacf(ar=c(0.6), ma=c(0.95), lag.max=100),xlim=c(1,40),ylab="r",type="h", main="ACF for ARMA(1,1)
phi=0.6, theta=0.95") ; abline(h=0)
```

5.2 Mixed Autoregressive –Moving Average models: ARMA(p, q)

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

Corresponding polynomials:

$\phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p$

$\theta(z) = 1 + \theta_1 z + \ldots + \theta_q z^q$

Notice signs of θ 's and ϕ's

ARMA(p, q) model using polynomials: $\phi(B) X_t = \theta(B) Z_t$

ARMA (p, q) is stationary, causal, has MA(∞) representation if

$\phi(z) \neq 0$ for all $|z| \leq 1$

equivalent: **the roots of the polynomial $\phi(z)$ lie outside of the unit circle**

ARMA(p, q) is invertible if polynomial $\theta(z) \neq 0$ for all $|z| \leq 1$

equivalent: **the roots of the polynomial $\theta(z)$ lie outside of the unit circle**

Principle of Parsimony – Economy of coefficients:
All stationary and invertible models can be written as MA(∞) or AR (∞)
Use the most parsimonious model (with the least coefficients).

Main Points to Take from Lecture 5

Autoregressive Models AR(p):

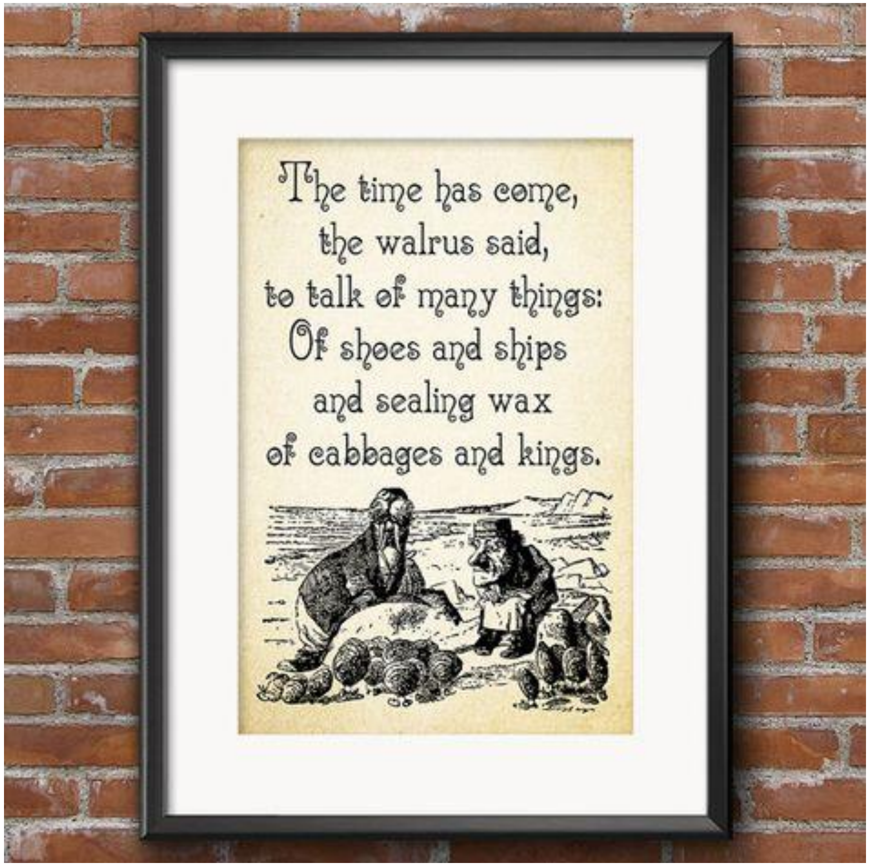
- Model equation: $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + Z_t$
  also $X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \ldots - \phi_p X_{t-p} = Z_t$
  also ϕ (B) X_t = Z_t
- AR models are always invertible.
- *AR models are stationary when they have MA(∞) representation.*
  Restriction on coefficients: |z*| > 1 for z* such that ϕ (z*) = 0.
- ACVF and ACF of AR(p) are obtained from the Yule Walker equations .
- Example of Yule – Walker equations are their solution for AR(2).

Mixed Models ARMA (1, 1) and ARMA (p, q):

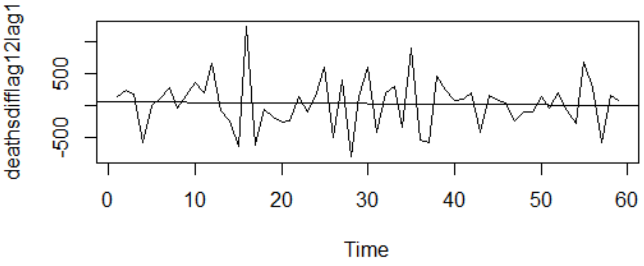
- Model equation: $X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}$
  or ϕ (B) X_t = θ (B) Z_t with $\phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p$, $\theta(z) = 1 + \theta_1 z + \ldots + \theta_q z^q$
- *If ϕ (z) $\neq 0$* for all $|z| \leq 1$ *, then* ARMA (p, q) is stationary, causal, and has MA(∞) representation $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$.
- If θ (z) $\neq 0$ for all $|z| \leq 1$, then ARMA(p, q) is invertible.
- Formulas for ARMA(1,1) for ψ_j's and ACVF.

End of Lecture 5

Welcome to Lecture 6: PACF. Non-stationarity. Differencing. Transformations. Seasonality. ARIMA and SARIMA Models



Data:



Models: **AR(p), MA(q), ARMA(p, q)**

Tools of Model Identification:

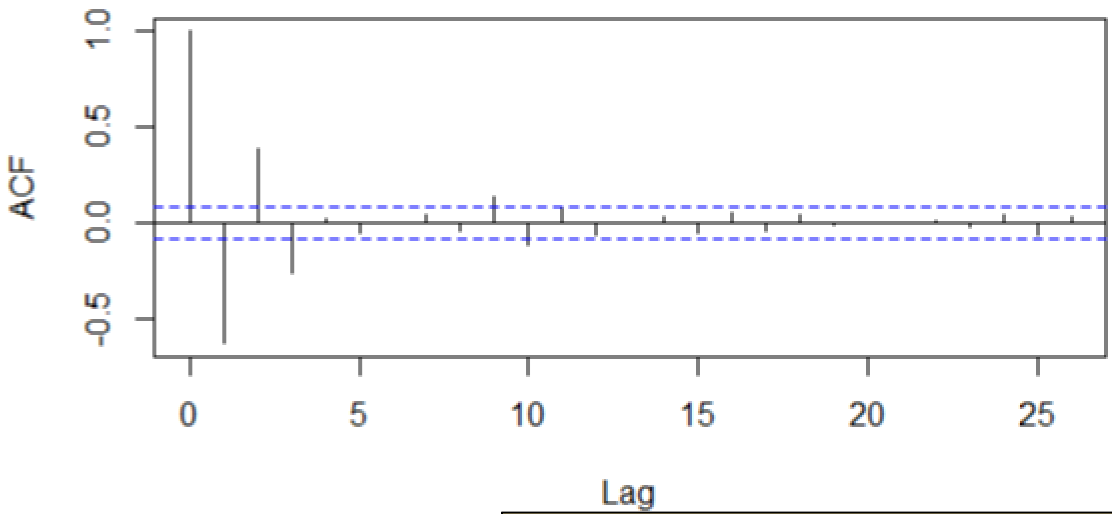
ACF, PACF

Outline of Lecture 6 of Week 3 Lectures

Lecture 6: PACF. Non-stationarity. Differencing. Seasonality. ARIMA Models. Transformations.

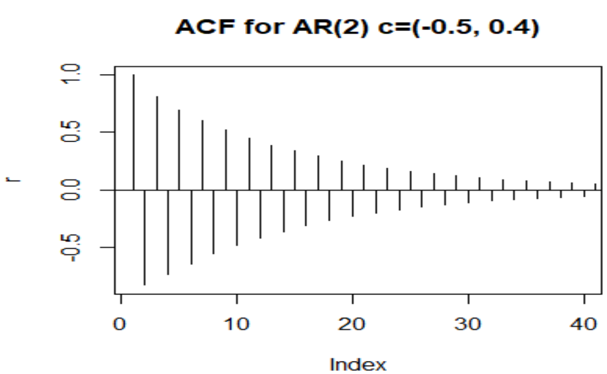
PACF: Identification of Parameter p in AR(p)

To identify q in MA (q), plot sample ACF and take q = max{k: $\widehat{\rho}_X(k) \neq 0$}



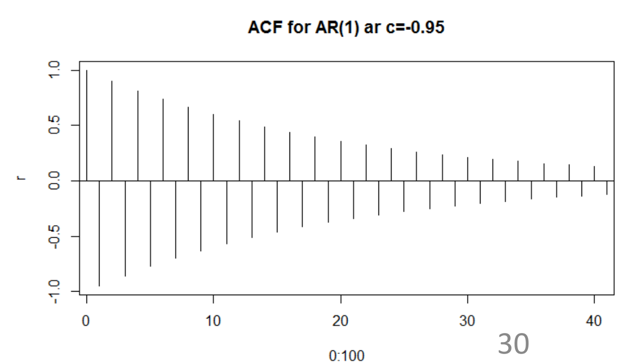
What is the order of MA q = ?

ACF is not very helpful to identify p for AR(p): ACF might look the same for different p.



ACF for AR(2)

ACF for AR(1)

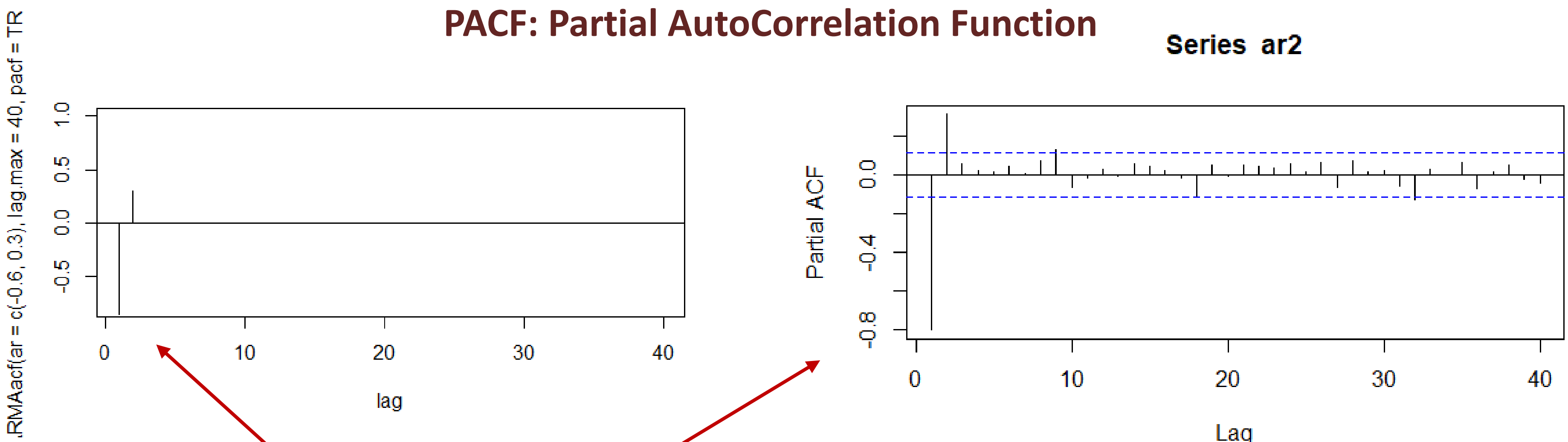


6. PACF: Identification of Parameter p in AR(p)

To identify p for AR(p), plot sample PACF $\hat{\alpha}(n) \equiv \hat{\phi}_{nn}$ and take
$p = \max\{ n: \hat{\alpha}(n) \neq 0\}$

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

PACF: Partial AutoCorrelation Function



Theoretical and Sample PACF for AR(2): $X_t = -0.6X_{t-1} + 0.3\,X_{t-2} + Z_t$

Partial Autocorrelation Function or PACF: Definition

Assume that data comes from AR process with unknown order n:

$X_t = \phi_{n1} X_{t-1} + \phi_{n2} X_{t-2} + \ldots + \phi_{nn} X_{t-n} + Z_t$ **where $Z_t \sim$ WN $(0, \sigma_Z^2)$**

Write Yule-Walker equations (§4.3 of Lecture Notes; slide 12 of this set)

$\rho_X(1) = \phi_{n1} + \phi_{n2}\,\rho_X(1) + \ldots + \phi_{nn}\rho_X(n-1)$;
$\rho_X(2) = \phi_{n1}\,\rho_X(1) + \phi_{n2} + \ldots + \phi_{nn}\rho_X(n-2)$;
.....
$\rho_X(n) = \phi_{n1}\,\rho_X(n-1) + \phi_{n2}\,\rho_X(n-2) + \ldots + \phi_{nn}$;

In matrix form:

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

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

Example: PACF for lags 1 and 2; general model.

Start with general Yule-Walker equations:

$\phi_{n1} \quad + \quad \phi_{n2}\, \rho_X(1) + \ldots + \phi_{nn}\rho_X(n-1) \qquad = \rho_X(1);$

$\phi_{n1}\, \rho_X(1) \ + \quad \phi_{n2} \quad + \ldots + \phi_{nn}\rho_X(n-2) \qquad = \rho_X(2);$

.....

$\phi_{n1}\, \rho_X(n-1) \ + \phi_{n2}\, \rho_X(n-2) + \ldots + \phi_{nn} \quad = \rho_X(n).$

In matrix form:

$R_n\, \underline{\phi}_n = \underline{\rho}_n;$

$\underline{\phi}_n = R_n^{-1}\, \underline{\rho}_n$

For n=1, we have one equation: $\phi_{11} = \rho_X(1)$,

that is, $\alpha(1) = \phi_{11} = \rho_X(1)$.

For n=2, we have two equation:

$\phi_{21} + \phi_{22}\, \rho_X(1) = \rho_X(1);$

$\phi_{21}\, \rho_X(1) \ + \phi_{22} = \rho_X(2)$. **Then,** $\alpha(2) = \phi_{22} = \dfrac{\rho_X(2) - (\rho_X(1))^2}{1 - (\rho_X(1))^2}$

For AR(p) model, $\phi_{pp} = \alpha(p)$ **is the last non-zero coefficient:**

$X_t = \phi_{n1}X_{t-1} + \phi_{n2}\, X_{t-2} + \ \ldots + \phi_{nn}\, X_{t-n} + Z_t;\ \phi_{pk} = 0$ for $k > p$ for AR(p).

Example 6.1: Find PACF for MA(1)

General Yule-Walker equations can be written for any process:

$$\phi_{n1} + \phi_{n2}\,\rho_X(1) + \ldots + \phi_{nn}\rho_X(n-1) = \rho_X(1)\,;$$
$$\phi_{n1}\,\rho_X(1) + \phi_{n2} + \ldots + \phi_{nn}\rho_X(n-2) = \rho_X(2)\,;$$
$$\ldots\ldots$$
$$\phi_{n1}\,\rho_X(n-1) + \phi_{n2}\,\rho_X(n-2) + \ldots + \phi_{nn} = \rho_X(n).$$

In matrix form:

$$R_n\,\underline{\phi}_n = \underline{\rho}_n\,;$$

$$\underline{\phi}_n = R_n^{-1}\,\underline{\rho}_n$$

For MA(1) model:

ACF is found in Example 3.2 of Week 1

For n=1, $\alpha(1) = \phi_{11} = \rho_X(1) = \theta_1/(1+\theta_1^2)$.

For n=2, $\alpha(2) = \phi_{22} = \dfrac{\rho_X(2) - (\rho_X(1))^2}{1-(\rho_X(1))^2} = \dfrac{0 - (\rho_X(1))^2}{1-(\rho_X(1))^2}$

$$= -\,\theta_1^2/(1+\theta_1^2+\theta_1^4)\,.$$

$$\alpha(k) = \phi_{kk} = (-1)^{(k+1)}\,\theta_1^k/(1+\theta_1^2+\ldots+\theta_1^{2k})\,,\ k>0.$$

Note: <u>for MA(1), PACF is always non-zero</u>.

Check your Understanding!

A stationary ARMA(p, q) model is known to have zero autocorrelations at all lags $k > 3$ and nonzero partial autocorrelation at lag one.
Determine (p, q). Explain your answer.

Make sure to be precise in your conclusions! Here is some help:

… zero autocorrelations at all lags $k>3$ …

--- We have $\rho_X(k) = 0$ for lags $k > 3$, i.e., $k = 4, 5, \ldots$, but …
Do we know the values of $\rho_X(k)$ for lags $k = 1, 2$ and 3?
Do these acfs have to be non-zero? Can be zeros?

… nonzero partial autocorrelation at lag one …

--- With exception of WN, all models might have $\phi_{11} \neq 0$, so that only WN is excluded by this sentence.

Check your answer of the next slide

Check your Understanding!

A stationary ARMA(p, q) model is known to have zero autocorrelations at all lags k > 3 and nonzero partial autocorrelation at lag one. Determine (p, q). Explain your answer.

Make sure to be precise in your conclusions! Here is some help:

… zero autocorrelations at all lags k>3 …

--- We have $\rho_X(k) = 0$ for lags k > 3, i.e., k= 4, 5, …, but do we know the values of $\rho_X(k)$ for lags k =1, 2 and 3?
Do these acfs have to be non-zero? Can be zeros?

… nonzero partial autocorrelation at lag one …

--- With exception of WN, all MA models might have $\phi_{11} \neq 0$, so that only WN is excluded by this sentence.

Zero autocorrelations at all lags k > 3 signifies that

--- it can only be pure MA(q) model with possible values of q, q = 0, 1, 2, 3.

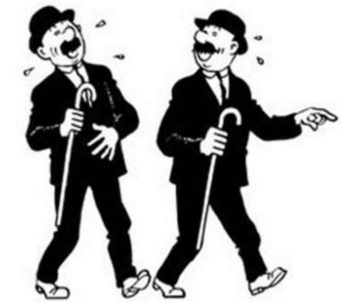
Nonzero partial autocorrelation at lag one

--- excludes q = 0, that is, excludes the White Noise model.

Conclude: Models MA(1), MA(2) and MA(3) are possible depending on the values of $\rho_X(k)$ for k = 1, 2 and 3.

For example, if $\rho_X(3) = 0$, but $\rho_X(2) \neq 0$, then this is a MA(2).

Math and Stats require <u>precision</u> so much that mathematicians laugh at themselves …



A Math Joke:

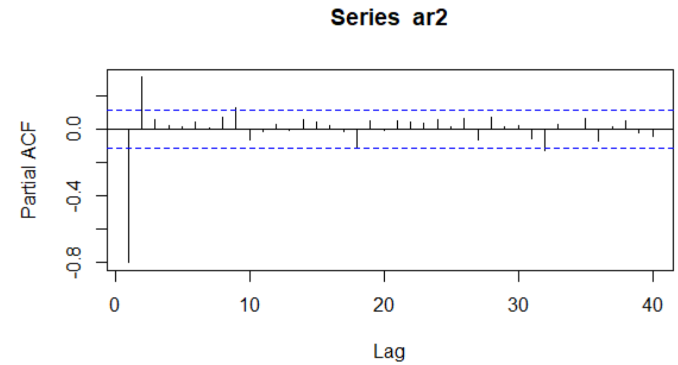
A mathematician, a physicist, and an engineer were travelling through Scotland when they saw a black sheep through the window of the train.

- **"Aha," says the engineer, "I see that Scottish sheep are black."**
- **"Hmm," says the physicist, "You mean that some Scottish sheep are black."**
- **"No," says the mathematician, "All we know is that there is at least one sheep in Scotland, and that at least one side of that one sheep is black!"**

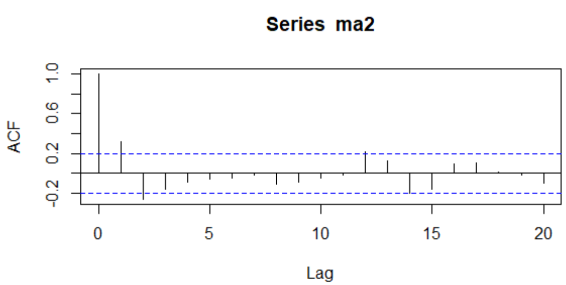
Summary: Identification of Parameters p in AR(p) and q in MA(q)

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)



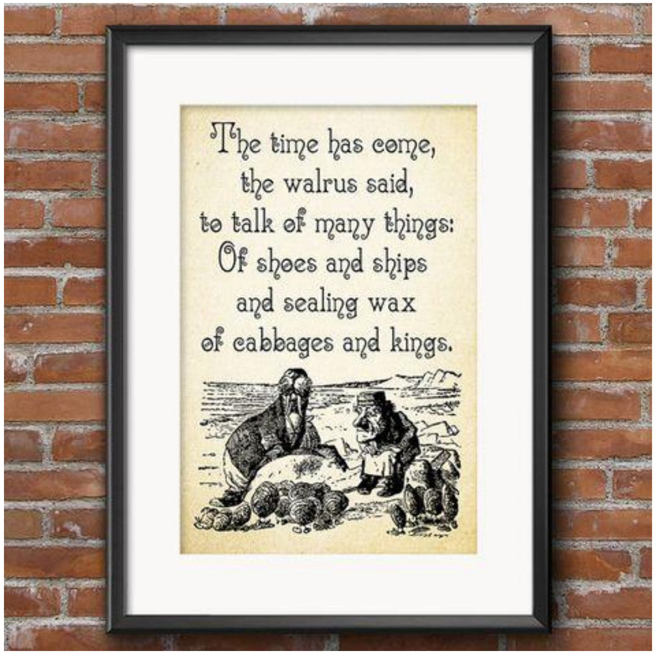
If sample ACF $\hat{\rho}(n) = 0$ for all $n > q$, but $\hat{\rho}(q) \neq 0$, then the data corresponds to MA process of order q.



ACF for MA(q)

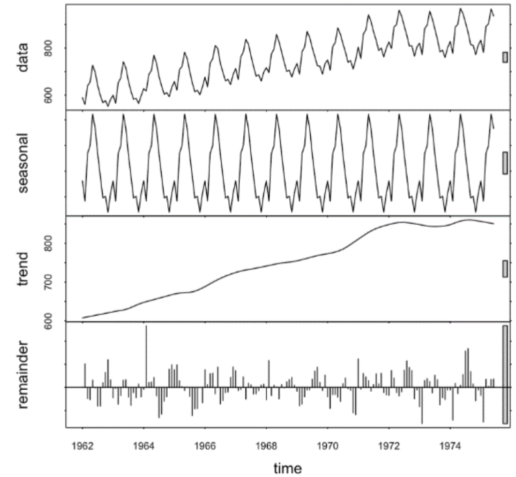
End of Part 1

To find the value of PACF at lag k, $\alpha(k)$, write a system of k Yule-Walker equations, and solve for the last coefficient: $\alpha(k) = \phi_{kk}$



Data:

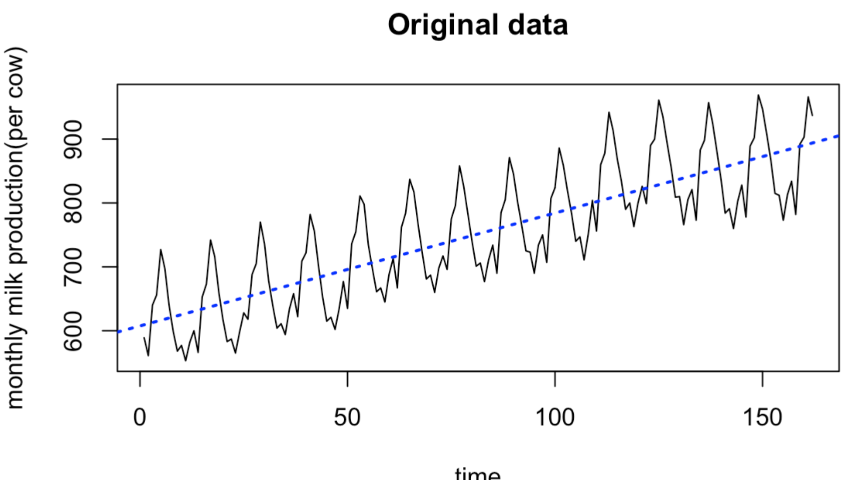
$$\boxed{X_t := m_t + s_t + S_t}$$

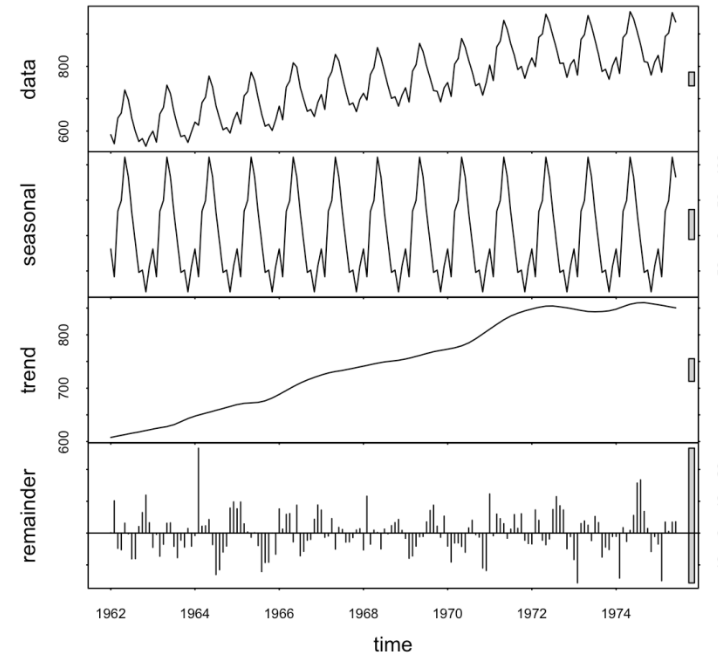


Tools to remove non-stationarity: differencing, ∇^d , ∇_s

Outline of Part II of Lecture 6: Non-stationarity. Differencing. ARIMA Models

7. Nonstationary Data : Trend and Seasonality





Classical Decomposition model:

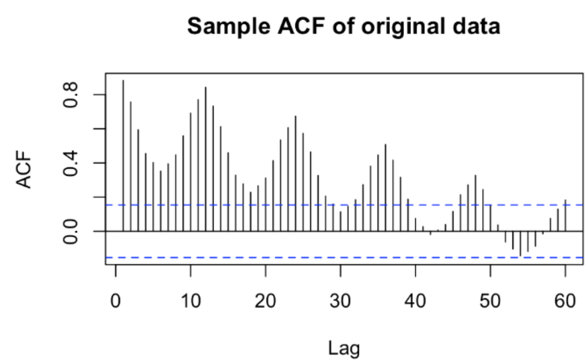
$X_t := m_t + s_t + S_t$

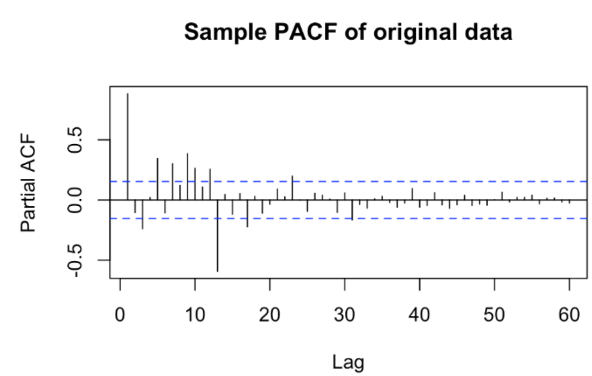
m_t = polynomial trend of order k;

s_t = seasonal component with period s:

$s_{t+s} = s_t$ or $\sum_{j=1}^{s} s_j = 0$

S_t = stationary process;





7. 1 Operator ∇ (NABLA). Differencing at lag 1

Before we proceed, we need
More new tools: OPERATOR ∇ (nabla)
Operation: Differencing

Differencing at lag 1:

$\nabla X_t \coloneqq X_t - X_{t-1}$ = (1 – B) X_t -- process X_t differenced at lag 1;

$\nabla^2 X_t \coloneqq \nabla(\nabla X_t) = \nabla X_t - \nabla X_{t-1} = X_t - 2X_{t-1} + X_{t-2}$ = (1 – B)^2 X_t

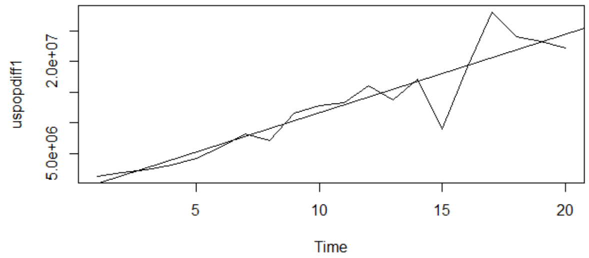
$\nabla^d X_t \coloneqq$ (1 – B)^d X_t – process X_t differenced d times at lag 1

Origin of the name of ∇ –a bit of history: https://en.wikipedia.org/wiki/Nevel_(instrument)
The nevel or nebel (Hebrew: נֵבֶל *nêḇel*) was a stringed instrument used by the ancient Hebrew people.
The Greeks translated the name as *nabla* (νάβλα, "Phoenician harp").

Used to eliminate trend -- See next slides

7.1 Elimination of Trend by Differencing at lag 1

Homogeneous nonstationarity:
No fixed mean level – trend;
Almost stationary apart from trend



Eliminate trend by differencing at lag 1:

$\nabla X_t := X_t - X_{t-1}$ = (1 – B) X_t -- process X_t differenced at lag 1

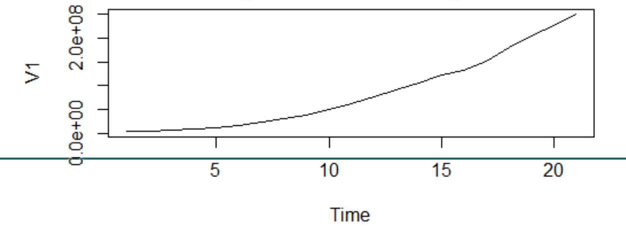
Example 7.1.1: Linear trend

X_t = *b t +* S_t *where* S_t *is a stationary process.*

Differencing once at lag 1 eliminates linear trend produces a stationary process:

$W_t = \nabla X_t = X_t - X_{t-1} = \{b\,t + S_t\} - \{b\,(t-1) + S_{t-1}\}$
$= b + \{S_t - S_{t-1}\}$
= ***stationary!***

7. 1 Elimination of Trend by Differencing at lag 1



Eliminate trend by differencing at lag 1:

$\nabla X_t \;:=\; X_t \;-\; X_{t-1}$ = (1 – B) X_t -- process X_t differenced at lag 1;

$\nabla^2 X_t \;:=\; \nabla(\nabla X_t\,) = \nabla\, X_t - \nabla X_{t-1} = \; X_t \;-\; 2X_{t-1} + X_{t-2} = (1 - \mathrm{B})^2\, X_t$

Example 7.1.2: Quadratic trend

X_t = b t^2 + S_t where S_t is a stationary process.

Differencing twice at lag 1 eliminates quadratic trend produces a stationary process:

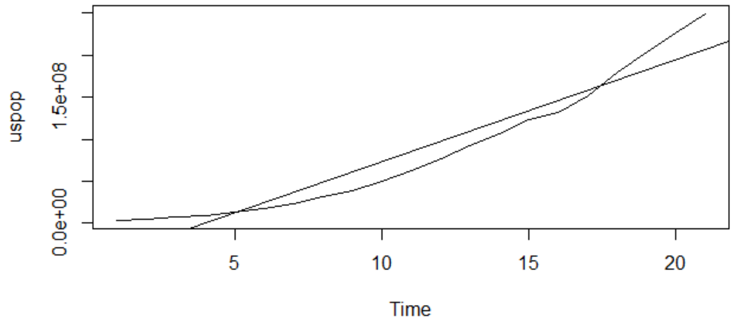
$$W_t = \nabla^2\, X_t = \; X_t \;-\; 2\, X_{t-1} + X_{t-2} =$$

$$= \{b\, t^2 + \; S_t\} - 2\{b\,(t-1)^2 + \; S_{t-1}\} + \{b\,(t-2)^2 + \; S_{t-2}\}$$

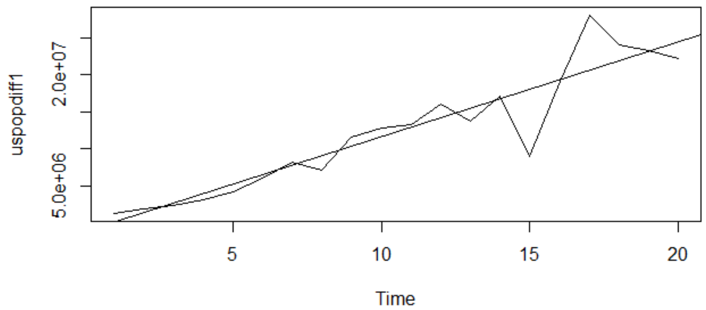
$$= b\{t^2 - 2\,(t-1)^2 + (t-2)^2\,\} + \{S_t - 2S_{t-1} + S_{t-2}\}$$

$$= 2b + \{S_t - 2S_{t-1} + S_{t-2}\} \;=\; \underline{\textit{stationary!}}$$

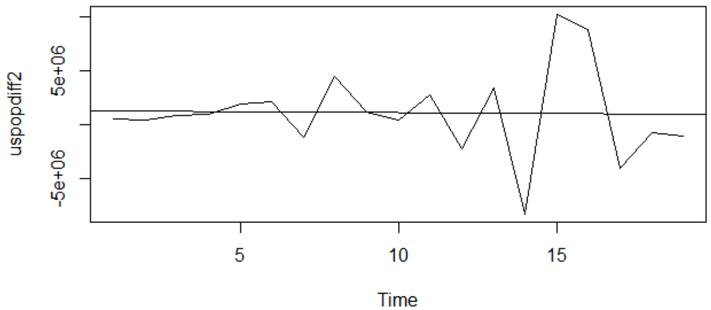
Example: USPOP data file and its differences at lag 1



US Population, 21 values,
Nonstationary: trend



First difference at lag 1 of USPOP
Still nonstationary: trend



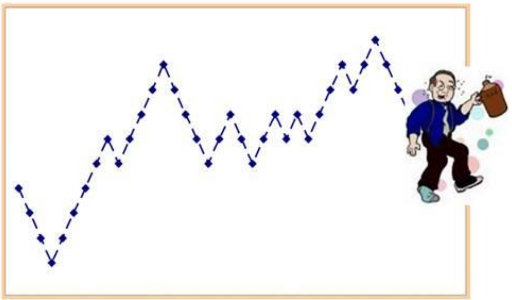
Second difference at lag 1 of USPOP
No trend.

Does it look stationary?

7. 1 Removing Non-Stationarity by Differencing

Differencing d times at lag 1 eliminates polynomial trend of order d.

Example 7.1.3: Random Walk Model



From Example 2.1.5 of week 1 (slide 55):

Let $Z_t \sim$ IID $(0, \sigma_Z^2)$ be I.I.D. WN, that is,
Z_t are i.i.d. r.v.s with mean zero and variance σ_Z^2.

Random Walk (RW) is defined as $X_t = Z_1 + \ldots + Z_t$.

Random Walk (RW) model can be written as $X_t = X_{t-1} + Z_t$.

Showed in 2.1.5: RW is non-stationary.

Difference at lag 1 produces a stationary process:

$$W_t = \nabla X_t = X_t - X_{t-1} = Z_t = WN = \textit{stationary!}$$

7. 2 Nonstationary Model: ARIMA(p,d,q)

Differencing d times at lag 1 eliminates polynomial trend of order d.

If the differenced (stationary) process follows ARMA(p,q), then the original (nonstationary) process is called ARIMA(p,d,q).

Mathematically,

if $W_t = \nabla^d X_t := (1-B)^d X_t$ **is stationary ARMA(p,q) :**

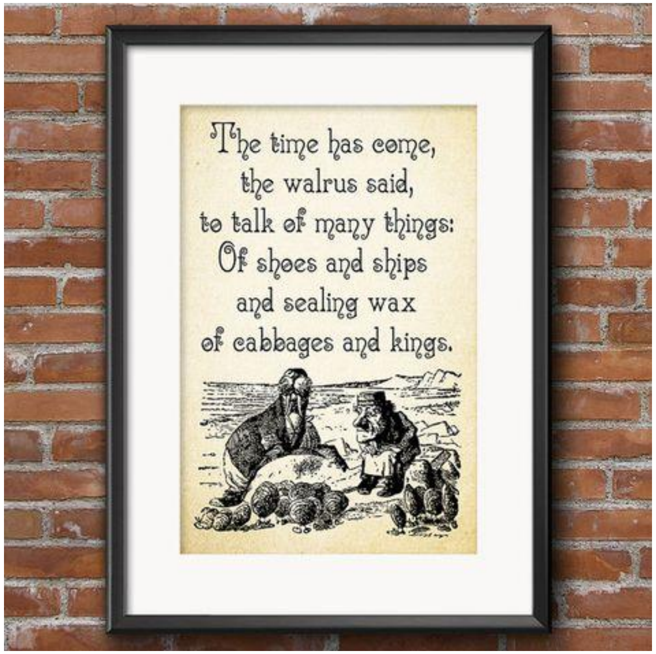
$$\phi(B)\, W_t = \theta(B)\, Z_t \;; \quad \phi(z) \neq 0 \;\; \& \;\; \theta(z) \neq 0 \;\; \text{for all } |z| \leq 1;$$

then ARIMA(p, d, q) model for X_t **is**

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

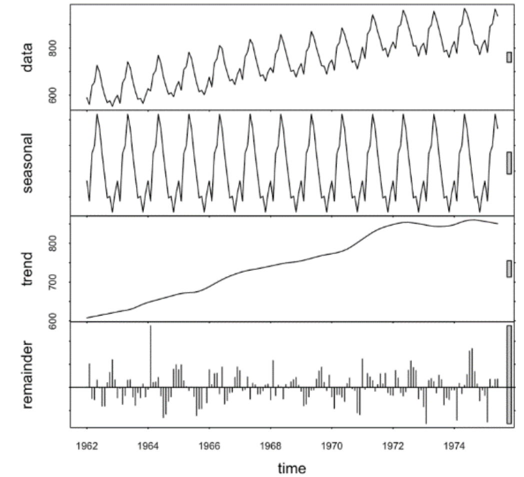
with $\phi^*(z) = \phi(z)\,(1-z)^d$ **that is,** $\phi^*(z)$ ***has a unit root of order d***

Next: learn to eliminate seasonality. Move to Part 3 of Lecture 6.



Data:

$$X_t := m_t + s_t + S_t$$



Tools to remove non-stationarity: differencing, ∇^d , ∇_s

Outline of Part III of Lecture 6:
Non-stationarity. Differencing. Seasonality. ARIMA Models

7. 3 Differencing at lag s to Remove Seasonality

Classical Decomposition model:

$$X_t := m_t + s_t + S_t$$

m_t = polynomial trend of order d;

s_t = seasonal component with period s: $s_{t+s} = s_t$ or $\sum_{j=1}^{s} s_j = 0$;

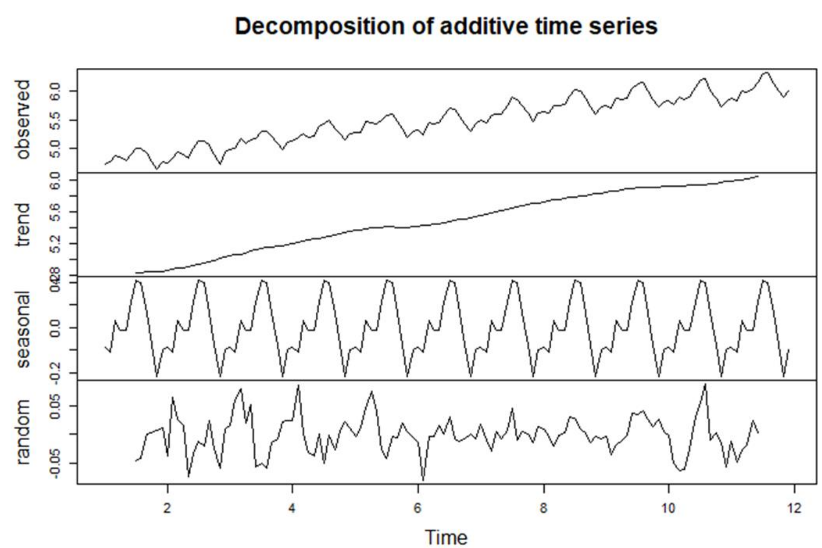
S_t = stationary process;

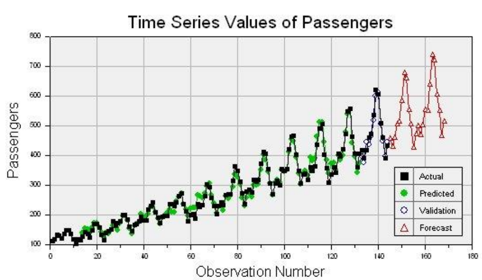
Differencing at lag s helps to remove seasonality:

$\nabla_s X_t := X_t - X_{t-s} = (1 - B^s)\, X_t$ -- process X_t differenced at lag s;

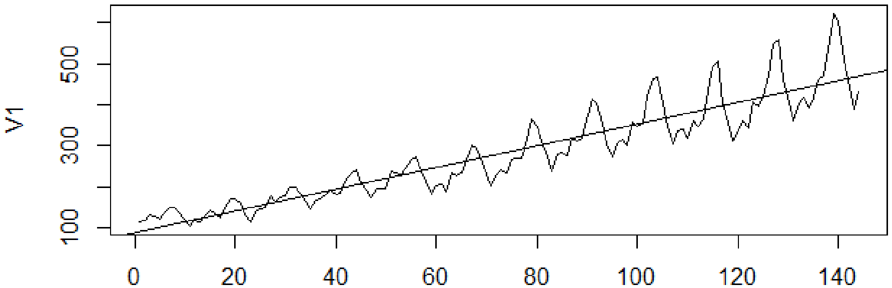
Differencing at lag s removes seasonality:

$$\nabla_s X_t := X_t - X_{t-s}$$
$$= (m_t - m_{t-s}) + (S_t - S_{t-s})$$
$$= \text{trend} + \text{Stationary process!}$$

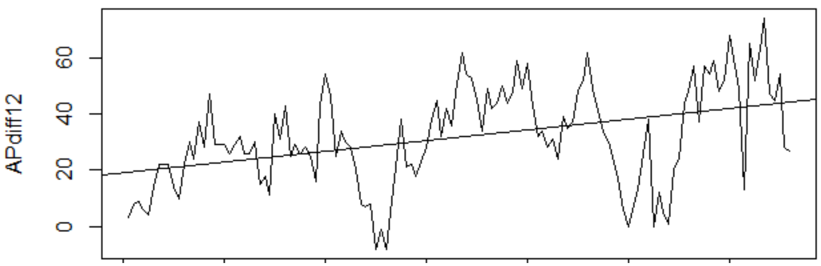




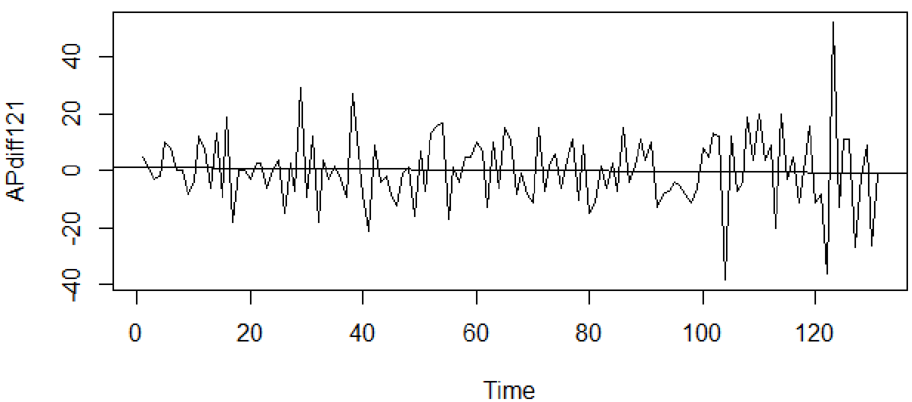
Example: International Airline Passengers Data



Airline passengers data: highly non-stationary
Linear trend; seasonality



Difference at lag 12: Removed seasonality

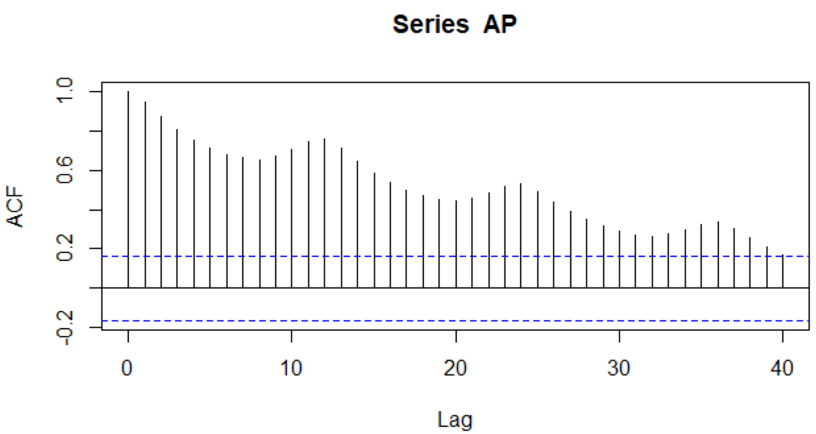


Additional differencing at lag 1 removed remaining trend!

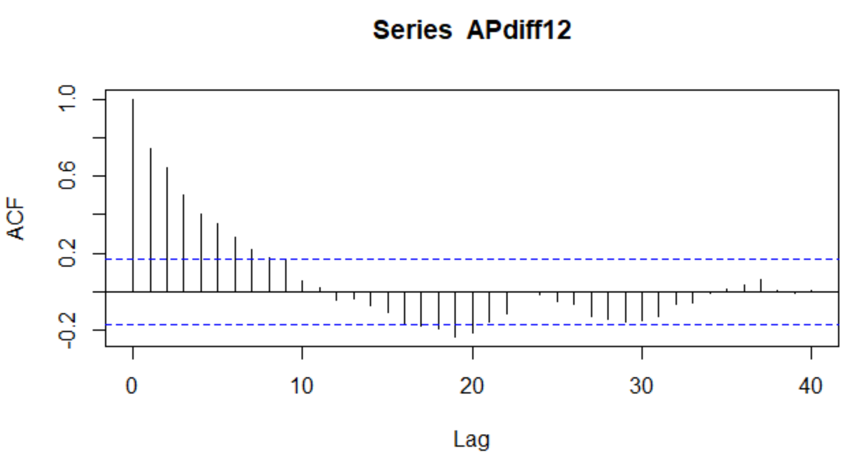
7. 4 More on Nonstationarity

(7.4.1) ARIMA(p, d, q) models are appropriate for data with trend

(7.4.2) Typical feature: ACF of nonstationary processes decay slowly



ACF for airline passengers data: highly non-stationary
Linear trend; seasonality

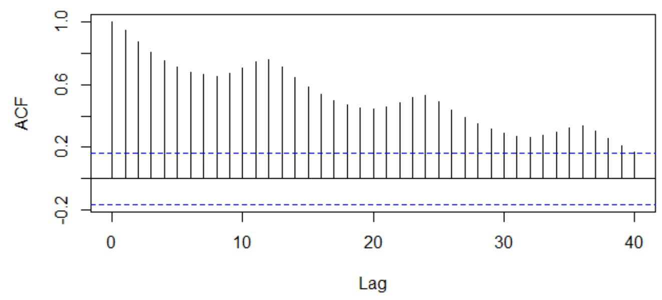


ACF for airline passengers data differenced at lag 12:
Non-seasonal; linear trend
Acf decays slowly

Check your Understanding

An actuary produces the following graph of sample acfs (called also correlogram) for vehicle accident severities over a 10-year period:

(lag in months).



You are also given the following three statements:

I. There is a positive trend in accident severity.

II. There is a negative trend in accident severity.

III. The accident severity data shows a seasonal pattern.

Determine which of the above statements can be concluded from the above graph:

A. I only

B. II only

C. III only

D. I, II, and III

E. The answer is not given by (A), (B), (C) or (D).

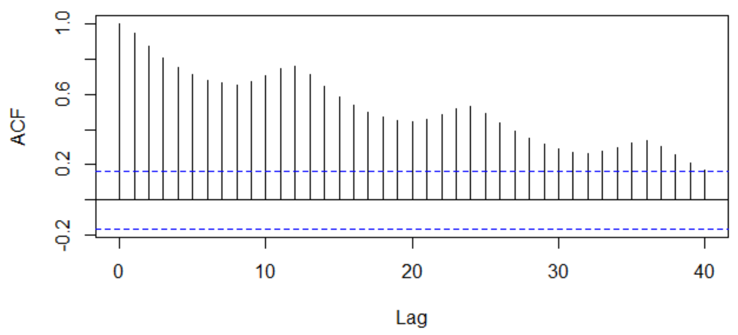
Accident severity is the amount of money that an insurance company will release for the policyholder in the event of an accident.
You do not need to use this information to answer the question above.

check your answer on the next slide

Check your Understanding

An actuary produces the following graph of sample acfs (called also correlogram) for vehicle accident severities over a 10-year period:

(lag in months).



You are also given the following three statements:

I. There is a positive trend in accident severity.

II. There is a negative trend in accident severity.

III. The accident severity data shows a seasonal pattern.

Determine which of the above statements can be concluded from the above graph:

A. I only

B. II only

C. III only

D. I, II, and III

E. The answer is not given by (A), (B), (C) or (D).

Answer: C. From the acf graph we see that the data is highly dependent and seasonal, because acfs are large and have seasonal variation.
We cannot, however, determine whether the data has a positive or negative trend from this graph.

Check your Understanding

You are given two models:

Model L: $Y_t = \beta_0 + \beta_1 t + Z_t$, where Z_t is a white noise process for t= 0, 1, 2, …

Model M: $Y_t = y_0 + \mu_c t + U_t$, where $C_t = Y_t - Y_{t-1}$, $U_t = \sum_{j=1}^{t} Z_j$, where Z_t is a white noise process for t= 0, 1, 2, …

Determine which of the above statements is/are true:

I. Model L is a linear trend in time model where the error component is not a random walk.

II. Model M is a random walk model where the error component of the model is also a random walk.

III. The comparison between Model L and Model M is not clear when the parameter μ_c =0.

A. I only

B. II only

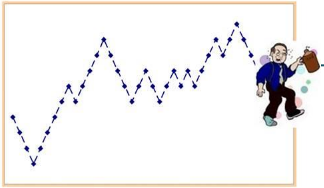
C. III only

D. I, II, and III

E. The correct answer is not given by (A), (B), (C) or (D).

check your answer on the next slide

Check your Understanding



You are given two models:

Model L: $Y_t = \beta_0 + \beta_1 t + Z_t$, where Z_t is a white noise process for t= 0, 1, 2, …

Model M: $Y_t = y_0 + \mu_c t + U_t$, where $C_t = Y_t - Y_{t-1}$, $U_t = \sum_{j=1}^{t} Z_j$, where Z_t is a white noise process for t= 0, 1, 2, …

Determine which of the above statements is/are true:

I. Model L is a linear trend in time model where the error component is not a random walk.

II. Model M is a random walk model where the error component of the model is also a random walk.

III. The comparison between Model L and Model M is not clear when the parameter μ_c =0.

A. I only

B. II only

C. III only

D. I, II, and III

E. The correct answer is not given by (A), (B), (C) or (D).

Answer: D.

Statement I is true: Model L has a linear trend $y = \beta_0 + \beta_1 t$ with an added random error Z_t being a white noise, not a random walk.

Statement II is true: Take a difference of Y in model M:

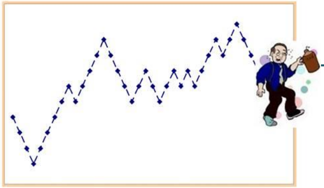
$\nabla Y_t = Y_t - Y_{t-1} = \mu_c\, (t-(t\text{-}1)) + (U_t - U_{t-1}) = \mu_c + Z_t$ = *WN with mean* μ_c. Compare with slide 45 on RW!

Statement III is true according to SOA. If μ_c =0 in model M, then Y_t is a RW starting at y_0;

It is hard to distinguish this RW from a WN sitting on a line, as in model L, if β_1 is not large.

-- see some simulations on the next slide.

Check your Understanding



You are given two models:

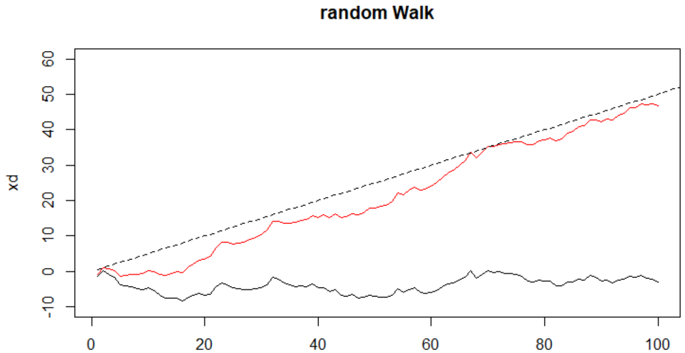
Model L: $Y_t = \beta_0 + \beta_1 t + Z_t$, where Z_t is a white noise process for t= 0, 1, 2, …

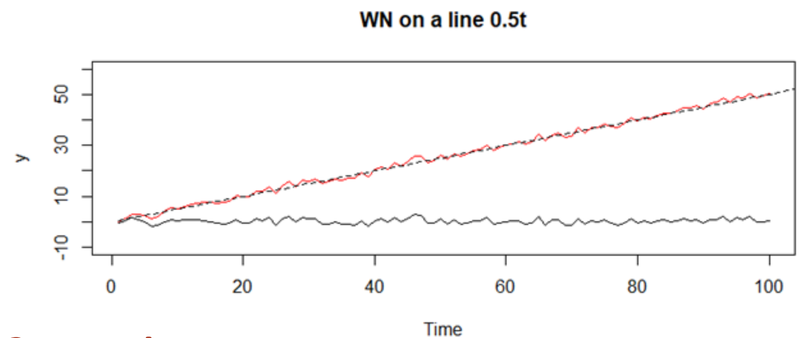
Model M: $Y_t = y_0 + \mu_c t + U_t$, where $C_t = Y_t - Y_{t-1}$, $U_t = \sum_{j=1}^{t} Z_j$, where Z_t is a white noise process

Model M is a RW with drift: $Y_t - Y_{t-1} = \mu_c + Z_t$ (see previous slide) or $Y_t = \mu_c + Y_{t-1} + Z_t$.

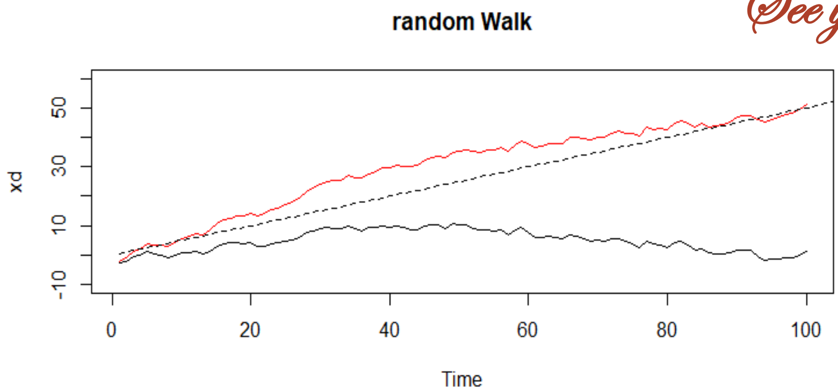
Observe its behavior for $\mu_c = 0.0$ (black) and $\mu_c = 0.5$ (red); $y_0 = 0$.

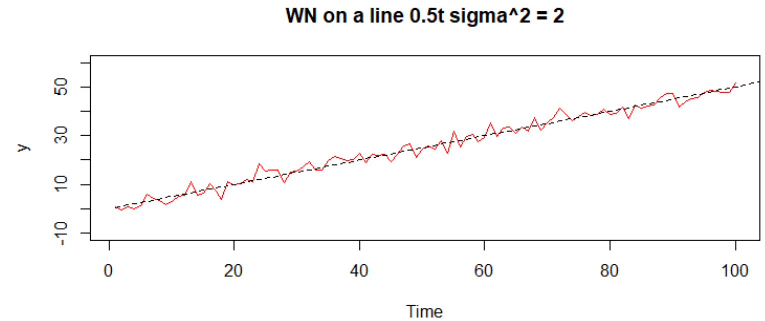
Cp., with the graph of WN on a line $Y_t = 0.5\,t + Z_t$.

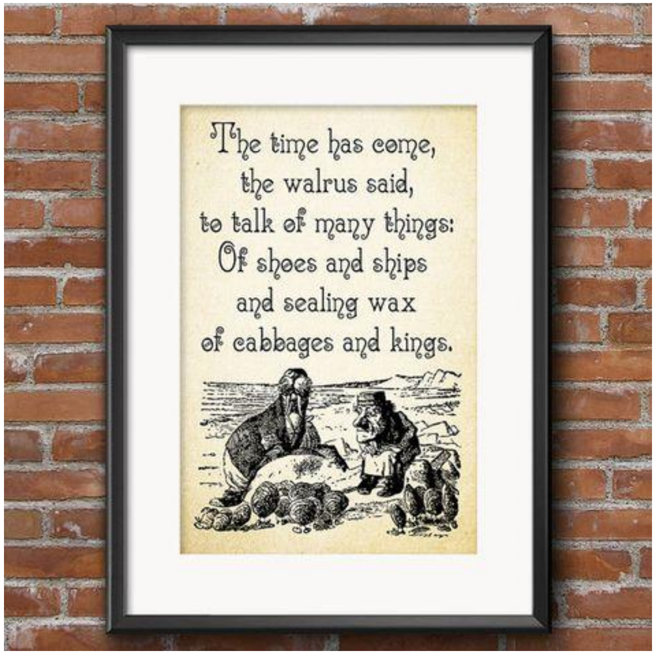




See you in Part IV !

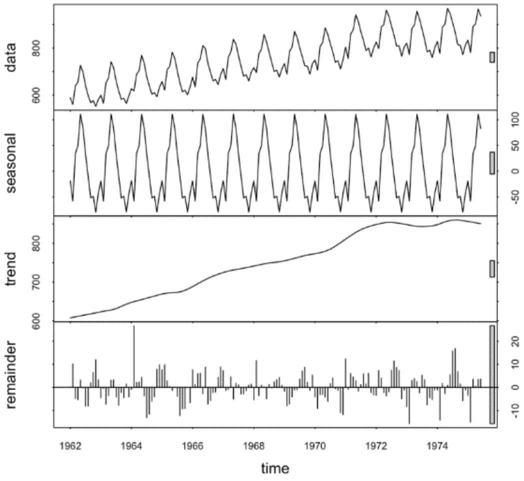






Data:

$$X_t := m_t + s_t + S_t$$



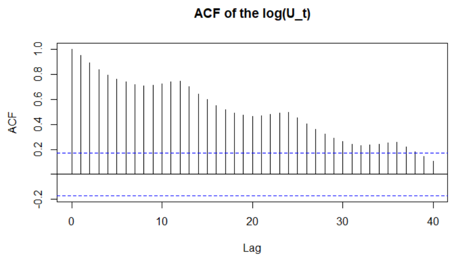
Tools to remove non-stationarity: differencing, transformation

Part IV of Lecture 6:
Trouble Shooting. Transformations. Summary and Code.

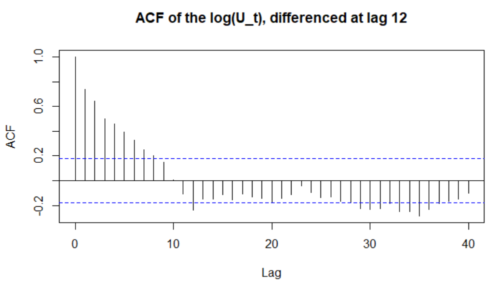
7. 4 ARIMA(p,d,q): Unit Roots, overdifferencing, etc.

How many times to difference? Did we take too many? Not enough? Tools?

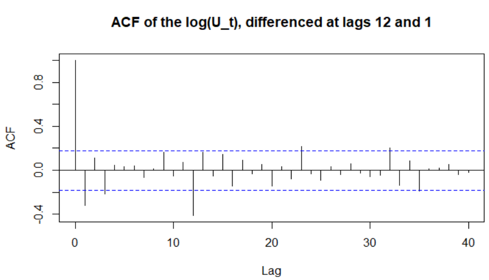
(7.4.1) ARIMA(p, d, q) models are appropriate for data with trend
(7.4.2) Typical feature: ACF of nonstationary processes decay slowly



ACF for highly non-stationary seasonal data
Linear trend; seasonality



ACF for the data after differencing at lag 12:
Non-seasonal; linear trend --Acf decays slowly



ACF after additional differencing at lag 1:
Looks like Stationary ACF

You think you did everything right, but got UNIT ROOTS in your model. Reasons?

7. 4 ARIMA(p,d,q): Unit Roots, overdifferencing, etc.

(7.4.3) Unit root in AR part: $\phi^* (B) X_t = \theta (B) Z_t$ with $\phi^*(1)=0$.

If $\phi^* (z) = \phi (z) (1 - z)^d$, $d \geq 1$, with $\phi (z) \neq 0$ for $|z| \leq 1$, then, differencing d times at lag 1 removes the unit root and produces a stationary process:

For X: $\phi^* (B) X_t = \phi(B) (1 - B)^d X_t = \theta (B) Z_t$, define $W_t = \nabla^d X_t = (1 - B)^d X_t$.

Then, W is stationary and follows ARMA model:

$$\phi (B) W_t = \theta (B) Z_t \text{ with } \phi (z) \neq 0 \text{ for all } |z| \leq 1. \Rightarrow \text{ stationary!}$$

Some algebra and application to AR(2) model with unit roots:

If $\phi^* (1) = 0$, then the polynomial $\phi^* (z)$ can be written as $\phi^* (z) = \phi (z) (1 - z)^d$, $d \geq 1$.

This fact is discussed in detail on slide # 5.

There might be several roots equal 1 as in this example:

Let $\phi^* (z) = 1-2.5z + 2z^2 - 0.5z^3$. Then, $\phi^* (1) = 0$ & $\phi^* (2) = 0$, that is, $\phi^* (z) = (1 - \frac{1}{2} z) (1-z)^2$.

Here, $d=2$, $\phi (z) = (1 - \frac{1}{2} z)$ has one root $z^* = 2 > 1$.

If X satisfies $(1-2.5B + 2B^2 - 0.5B^3) X_t = \theta (B) Z_t$, then $W_t = (1-B)^2 X_t$ satisfies a stationary model

$$(1 - \frac{1}{2} B) W_t = \theta (B) Z_t$$

7. 4 ARIMA(p,d,q): Unit Roots, overdifferencing, etc.

(7.4.4) Unit root in MA part means overdifferencing:

- **Let Y_t be a stationary and invertible ARMA:**

$$\phi(B) Y_t = \theta(B) Z_t \text{ with } \theta(z) \neq 0 \text{ for all } |z| \leq 1.$$

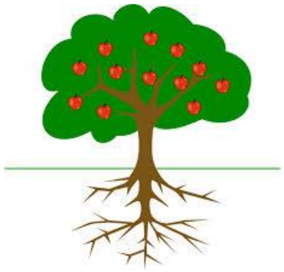
What happens if we difference it unnecessarily (overdifference)?

➢ ***Take $X_t = \nabla Y_t = (1 - B) Y_t$. Then***

$$\phi(B) X_t = \phi(B)(1 - B) Y_t = (1 - B) \phi(B) Y_t$$

$$= (1 - B)\, \theta(B) Z_t = \theta^*(B) Z_t$$

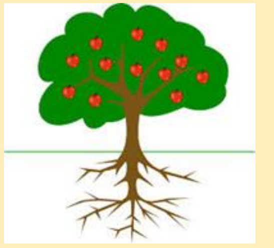
Conclude: By differencing an invertible ARMA model, we introduced the unit root in MA part making it not invertible.



Young Woman Picking Oranges by Berthe Morisot

7. 4 ARIMA(p,d,q): Unit Roots, overdifferencing, etc.

Practical Conclusion from (7.4.3) and (7.4.4):



Got units roots? Check your differencing!

Unit roots in AR part : try differencing at lag 1;

Unit roots in MA part: check overdifferencing

Additional tip on getting differencing right:

check change of variance!

See explanation in 7.4.5 and examples on the next slides.

7.4.5 ARIMA(p,d,q): Effect of Overdifferencing on Variance

(7.4.5): Overdifferencing increases variance:

Let $X_t = Z_t + \theta_1 Z_{t-1}$ = $(1 + \theta_1 B) Z_t$, $|\theta_1| < 1$, be MA(1), with
ACF: $\rho_X(0) = 1$; $\rho_X(k) = 0$ for $|k| > 1$, and $\gamma_X(0) = \sigma_Z^2 (1 + \theta_1^2)$.

Difference X: $W_t = \nabla X_t = (1 - B) X_t = (1 - B)(1 + \theta_1 B) Z_t$

$$= Z_t - (1 - \theta_1) Z_{t-1} - \theta_1 Z_{t-2}$$

Conclude: W_t **is MA(2) with** $\rho_W(2) \neq 0$**, and variance:**

$$\gamma_W(0) = \sigma_Z^2 \{1 + (1 - \theta_1)^2 + \theta_1^2\} = \sigma_Z^2 \{1 + \theta_1^2\} + (1 - \theta_1)^2$$

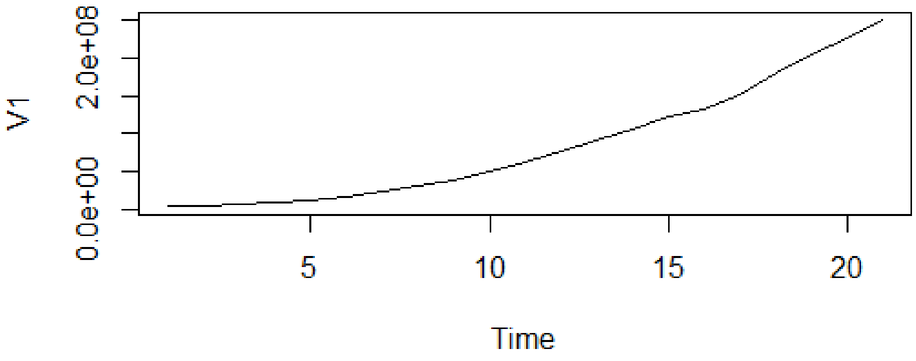
$$= \gamma_X(0) + \text{a positive term.}$$

Conclude: $\gamma_W(0) > \gamma_X(0)$, that is, Unnecessary differencing

-- increased variance!

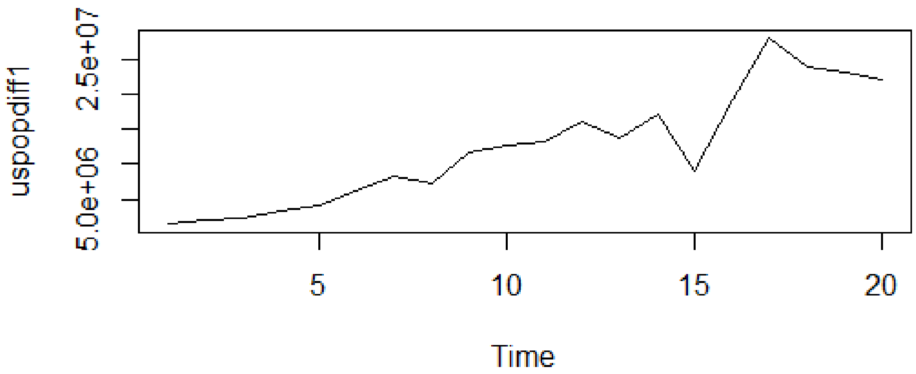
-- increased the lag of nonzero acfs!

USPOP data file and its differences



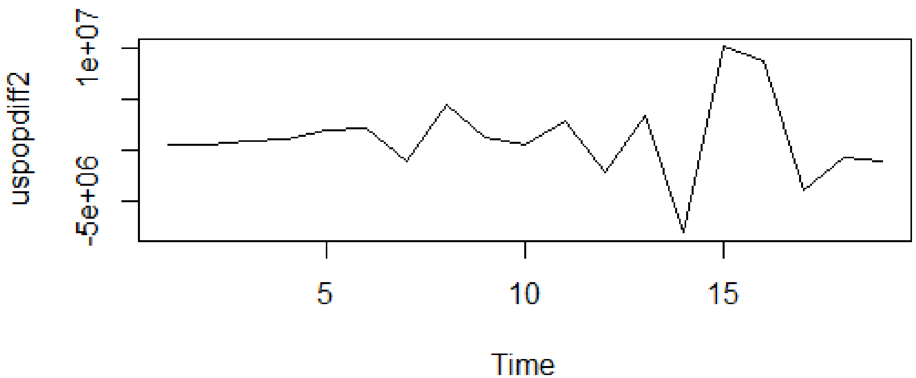
US Population, 21 values,

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



First difference of USPOP

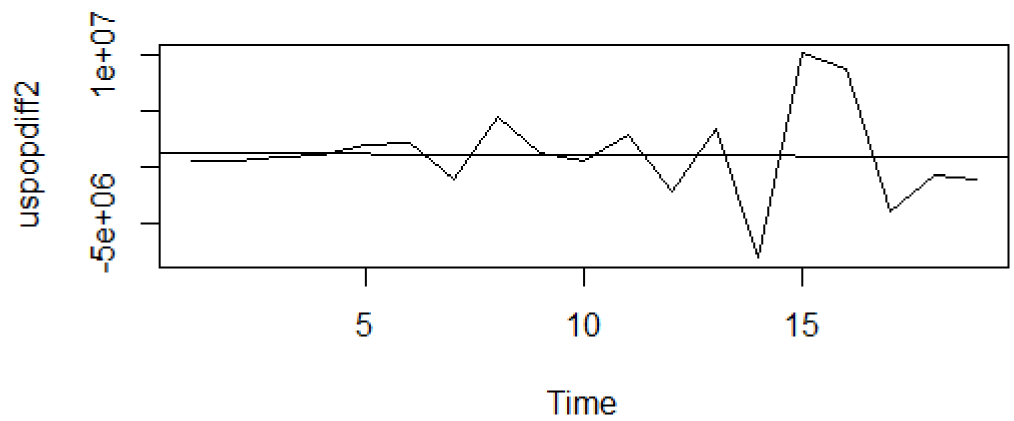
> uspopdiff1 <- diff(uspop,differences = 1);
> var(uspopdiff1)
[1] 6.597748e+13



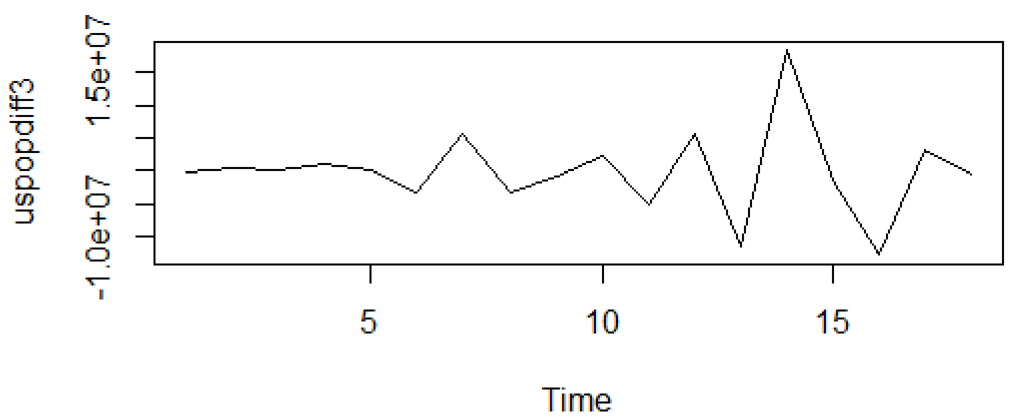
Second difference of USPOP;

>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

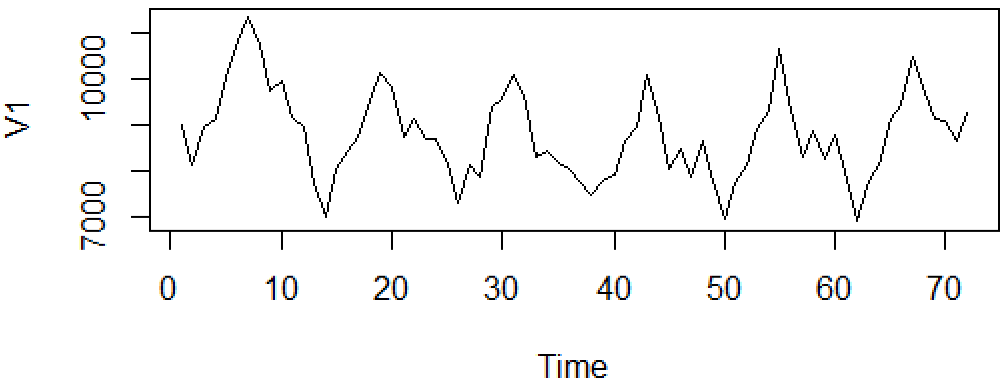


>var(uspopdiff3)
[1] 4.525136e+13

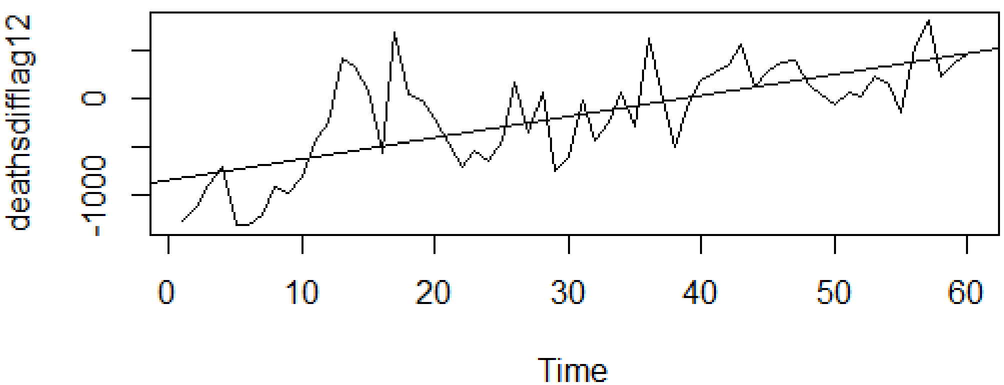
Note: Variance of the series decreased at differences 1 and 2 but increased when differencing the third time. What do you conclude?

(third differencing is unnecessary, overdifferencing)

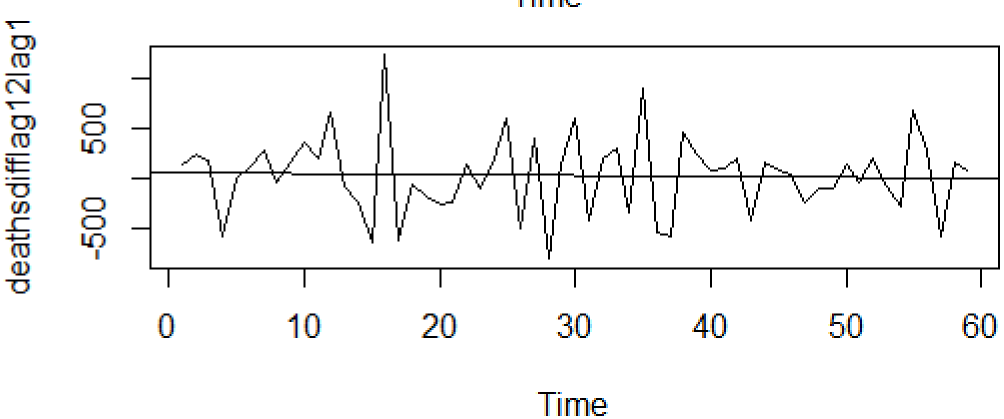
Accidental Deaths Data, differenced at lag 12 and then at lag 1



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

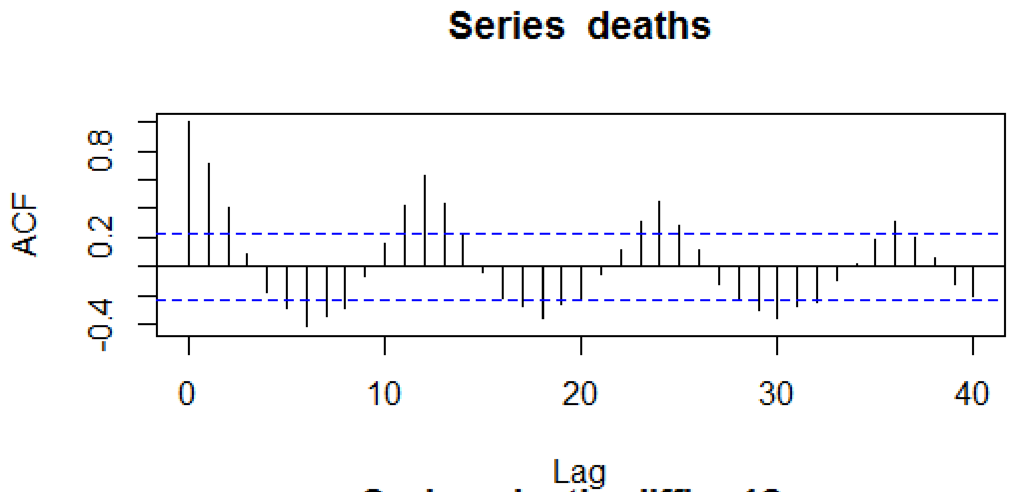


Difference at lag 12 to remove seasonality
Sample variance: 288,714.5

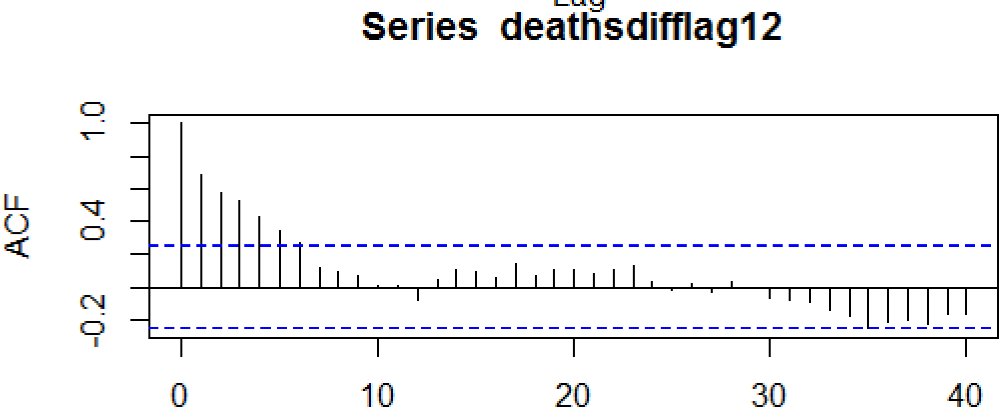


Differenced again at lag 1 to remove trend: $\nabla\nabla_{12} X_t$
Sample variance: 15,5301.9

ACF for Accidental Deaths Data and its differences



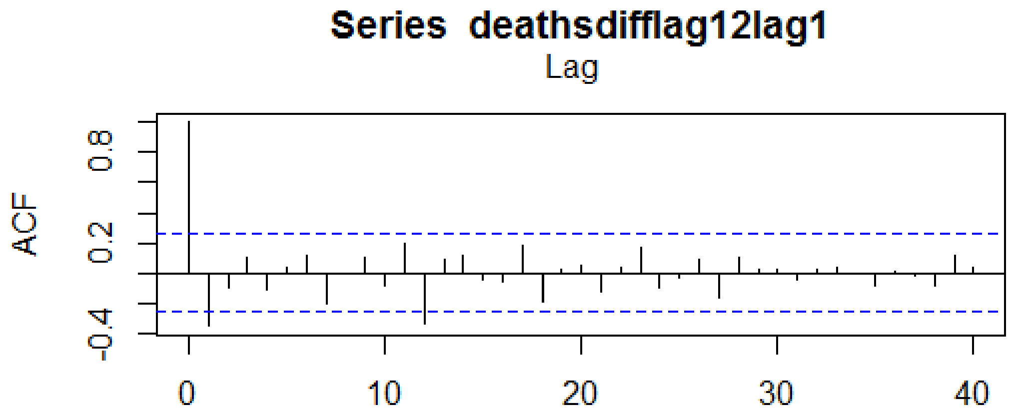
ACF for original data: the monthly accidental deaths, 1973-1978



ACF for difference at lag 12

No longer periodic at lag 12

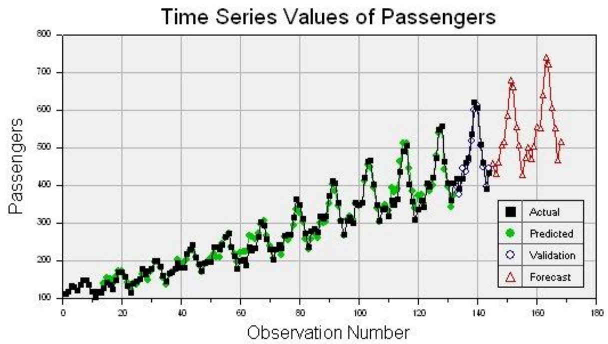
Slow decay



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

8. Transformations of Data: The Box-Cox. Log.

- **Reasons to use transformation:**
  - **to stabilize variance/seasonal effect;**
  - **to make data approximately normal**



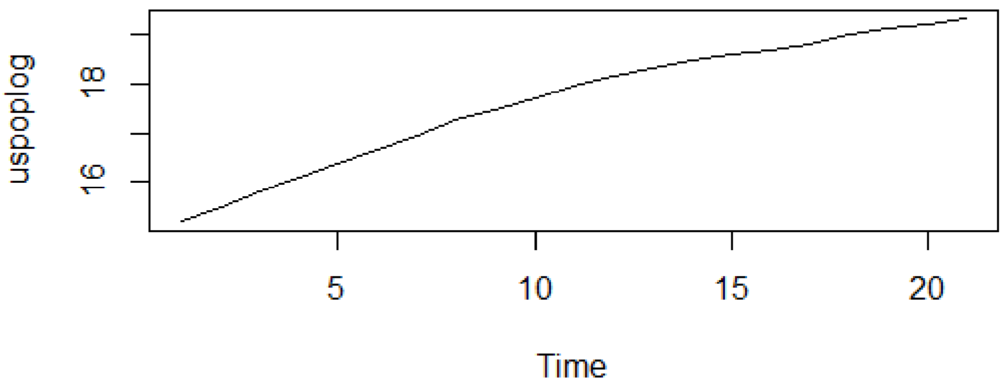
The Box-Cox Transform of original data U_t:

$$f_\lambda(U_t) = \begin{cases} \ln U_t & \text{if } U_t > 0,\ \lambda = 0; \\ \lambda^{-1}(U_t^\lambda - 1), & \text{if } U_t \geq 0,\ \lambda \neq 0 \end{cases}$$

Used most often:
Log transform: X_t = ln (original data);
Powers, e.g., square root: $X_t = \sqrt{original\ data}$

Shown in the §8 of Lecture Notes that $f_\lambda(x) \to f_0(x) = \ln x$ when $\lambda \to 0$.

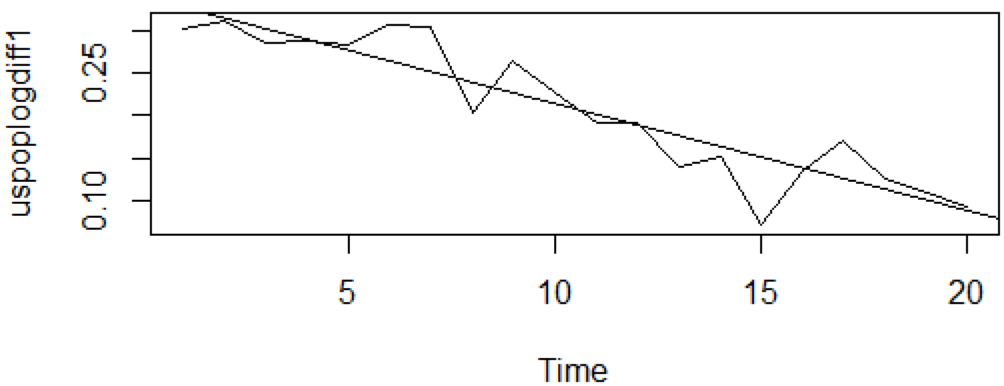
Log(USPOP) and its differences



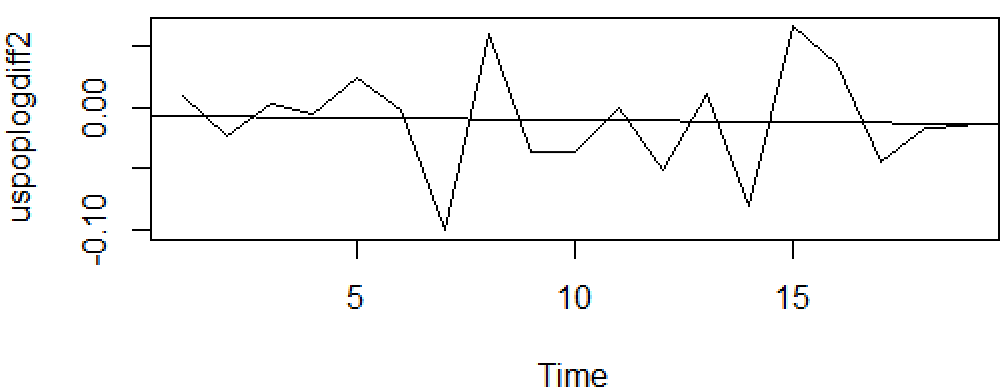
Plot of Log(USPOP)

> var(uspoplog)

[1] 1.716024



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

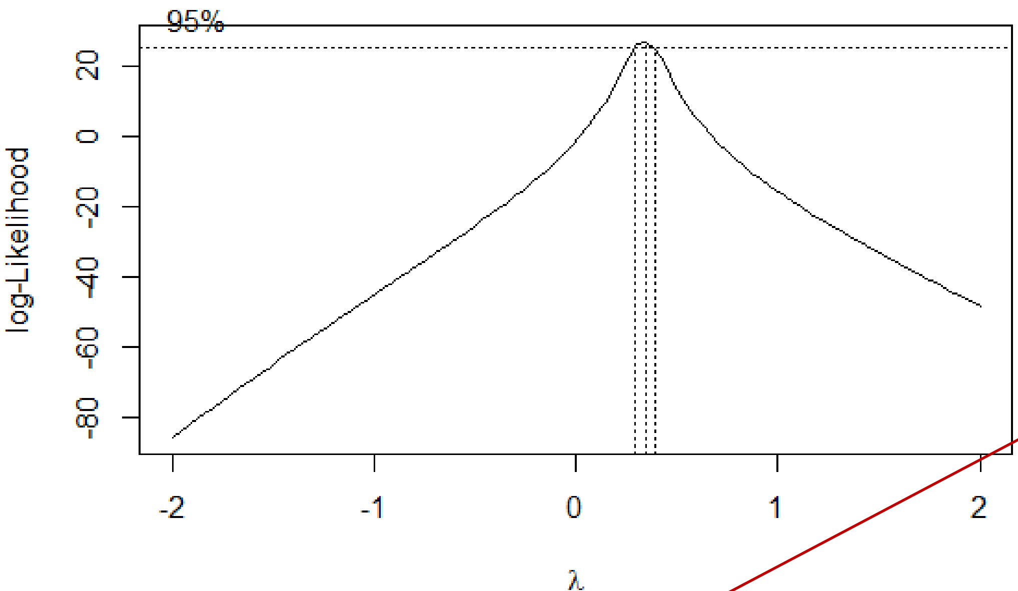


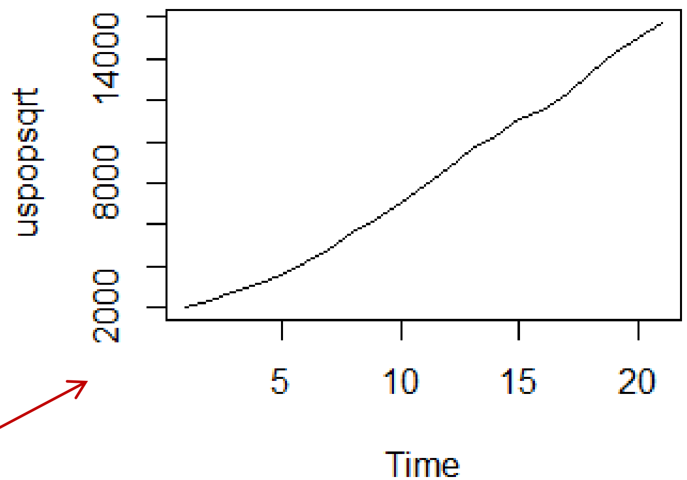
Second difference of log(USPOP); sample variance is 0.001799062

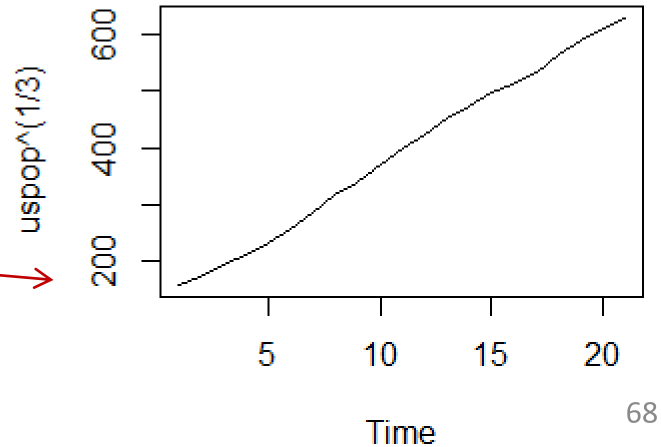
8. Box-Cox Transformation of USPOP

 # 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 and cubic root transformations
λ=1/2 could be in the confidence interval.
λ= 0 (log) not in the confidence interval.

Main Points to Take from Lecture 6

- PACF as a tool to identify order p of AR(p): definition & how it works
- Nonstationary Models. Classical Decomposition Model.
- Differencing, Operator ∇.
- Elimination of trend by differencing at lag 1 and

  of seasonality with period s by differencing at lag s
- ARIMA(p,d,q) model (when it arises and an equation)
- Choosing the right number of differencing:
  - Changes in variance
  - Unit root in AR part – difference again!
  - Unit root in MA part – unnecessary differencing!
- Transformations, Box-Cox transforms

Some simple R commands used to create previous slides

- **To find roots of the polynomial $\phi(z) = 1 - 1.3z + 0.7z^2$:**

```
> polyroot(c(1, -1.3, 0.7))
```

- **To plot 100 values of ARMA(1,1) process and its theoretical acf:**

```
plot.ts(arima.sim(model=list(ar=c(0.8), ma=c(0.3)), n = 100, sd = 1), main="ARMA(1,1), ar c=0.8, ma c=0.3")
plot(0:100,ARMAacf(ar=c(0.8), ma=c(0.3), lag.max=100),xlim=c(1,40),ylab="r",type="h", main="ACF for ARMA(1,1) ar c=0.8, ma c=0.3") ; abline(h=0)
```

- **To simulate 300 values of AR(2): $X_t = -0.6X_{t-1} + 0.3X_{t-2} + Z_t$ and plot 40 pacfs:**

```
ar2 <- arima.sim(model=list(ar=c(-0.6, 0.3)), n=300, sd=1)
pacf(ar2, 40)
```

- **To plot theoretical ACF for this AR(2)**

```
plot(ARMAacf(ar=c(-0.6, 0.3), lag.max=40), col="red", type="h", xlab="lag", ylim=c(-.8,1)); abline(h=0)
```

- **To plot theoretical PACF for this AR(2)**

```
plot(ARMAacf(ar=c(-0.6, 0.3), lag.max=40, pacf=TRUE), type="h", xlab="lag", ylim=c(-.8,1)); abline(h=0)
```

(continued)

Some simple R commands used to create previous slides

- **To plot US population data with regression line:**

```
uspop <- scan("uspop.txt")
plot.ts(uspop)
fit <- lm(uspop ~ as.numeric(1:length(uspop))); abline(fit)
```

- **To plot first difference at lag 1 of US population data with regression line:**

```
uspopdiff1 <- diff(uspop)
plot.ts(uspopdiff1)
abline(lm(uspopdiff1 ~ as.numeric(1:length(uspopdiff1))) )
```

- **To find variance:**

```
var(uspop)
```

- **To plot graph of airline data saved in the file pass.txt and its ifference at lag 12:**

```
AP=read.table("pass.txt")
plot.ts(AP)
APdiff12 <- diff(AP, lag=12, differences = 1)
plot.ts(APdiff12)
abline(lm(APdiff12 ~ as.numeric(1:length(APdiff12))) )
```

- **To choose parameter λ of the Box-Cox transformation for USPOP datset:**

```
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
```

Some simple R commands used to create previous slides

- **To plot RW with and w/o drift:**

```
set.seed(12)
#White noise
w=rnorm(100,0,1)
#Random walk without drift
x=cumsum(w)  #cummulative sum
#Random walk with drift 0.5
wd=w+0.5
xd=cumsum(wd)
plot.ts(xd,main="random Walk",col="red",ylim=c(-10,60))
lines(x)
lines(.5*(1:200), lty="dashed")
```

- **To plot WN on a line y_t = 0.5t + WN:**

```
y = c(0)
for(i in 1:100){y[i] = 0.5*i + rnorm(1,0,1)}
plot.ts(y, main="WN on a line 0.5t",col="red",ylim=c(-10,60))
lines(.5*(1:200), lty="dashed")
```