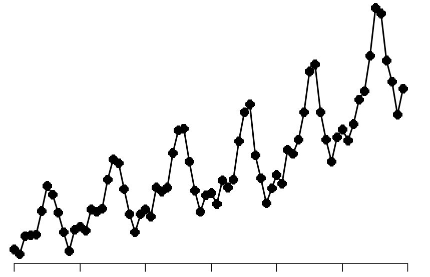


1. Time Series Fundamentals

STAD57 F19

Sotirios Damouras

Adminis-trivia

- *Instructor*: Sotirios Damouras
  - Pronounced s-aw-t-EE-r-ee-aw-s or Sam
- *Contact Info*:
  - *email*: sdamouras@utsc.utoronto.ca
  - *Office hours*: WE 15-17 @ IC456, or by appointment (email)
- *Course web page on Quercus* :
  - All course material (outline, lecture slides, assignments & solutions) posted on portal

Outline

- *Textbook*:
  - *Time Series Analysis and Its Applications, with R examples 4th Edition;* R.H. Shumway & D.S. Stoffer
    - Cover (parts of) §1-5, with extra topics if time permits
  - *Forecasting: Principles and Practice*; R.J. Hyndman & G. Athanasopoulos
- *Evaluation*:
  - Course project worth 20%
  - Midterm exam worth 30%
    - If you miss test/exam, must submit UTSC medical certificate to take *make-up* test/exam
  - Final exam, worth 50%

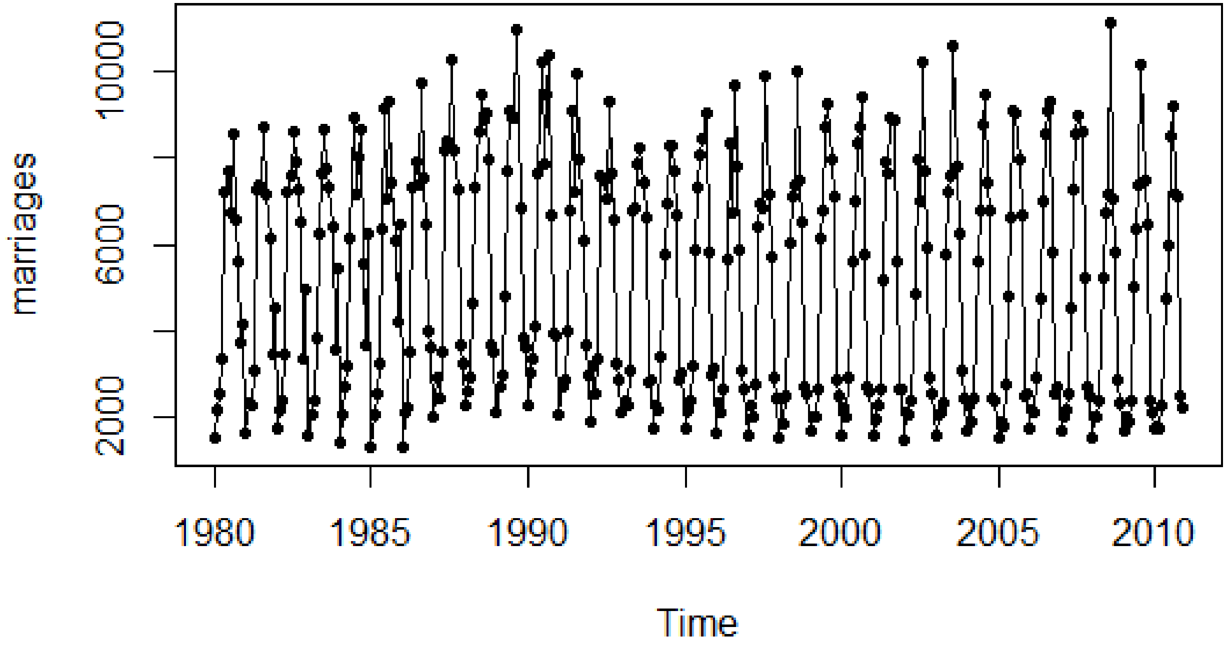
Time Series Data

- Data recorded *over time*, typically at regular intervals (daily, monthly, etc)
  - E.g. monthly number of marriages in Ontario (from Ontario open data http://www.ontario.ca/open-data)

Year	*Month*	*Count*
1980	Jan	1,533
1980	Feb	2,149
1980	Mar	2,571
1980	Apr	3,345
1980	May	7,220
1980	Jun	7,707
1980	Jul	6,744
⋮	⋮	⋮
2010	Dec	2,247

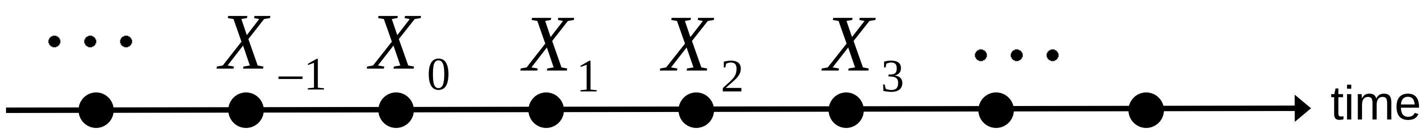
Time Series Plot

- Similar to scatter plot, where:
  - X-axis is time and Y-axis is data variable
  - points connected with line to show order



Time Series

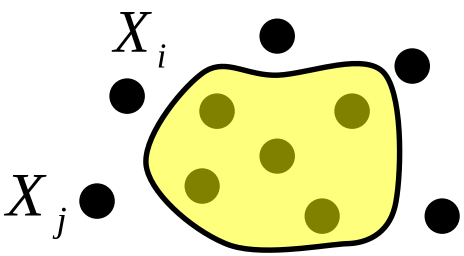
- Time Series (TS) formal definition: *ordered sequence of Random Variables* (RV)
  - Graphical representation: RVs are bullets ordered along a time line



  - Notation: $\{X_t\}_{t\in\mathbf{Z}} = \{\ldots, X_{-1}, X_0, X_1, X_2, \ldots\}$
  - Essentially, TS = stochastic process

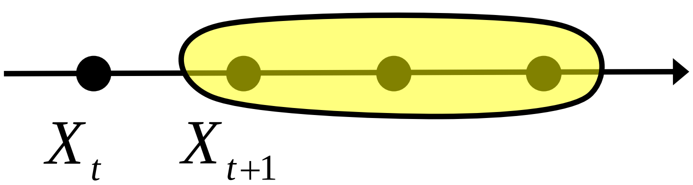
Time Series Analysis

- "Usual" Stats: randomly sample population



data $x_1, \ldots, x_n$
are independent

- Time Series: observe series over time

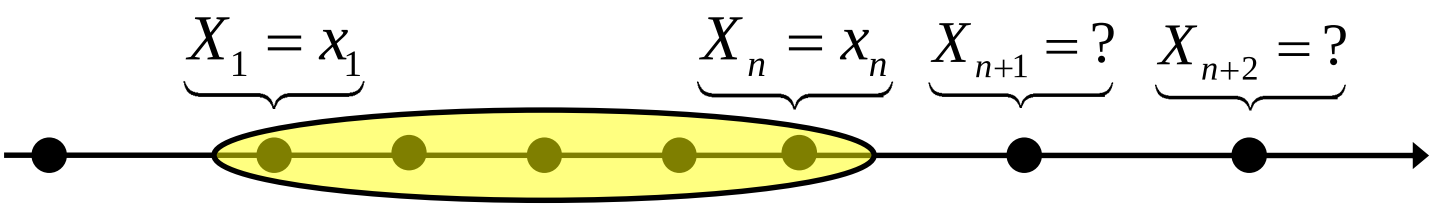


data $x_1, \ldots, x_n$
are *dependent*

Time Series Analysis

Two basic functions of TS analysis

- *Estimation*: Describe how RVs $\{X_t\}_{t\in\mathbf{Z}}$ are related based on data $x_1, x_2, \dots, x_n$
- *Prediction/Forecasting*: Predict future values of $X_{n+1}, X_{n+2}, \dots$ given $x_1, x_2, \dots, x_n$



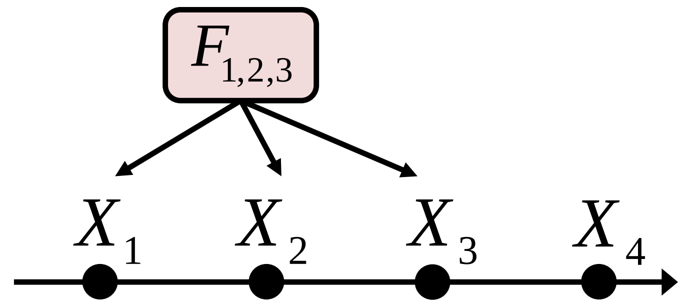
Both require a model for the TS $\{X_t\}_{t\in\mathbf{Z}}$

Time Series Modeling

- Ideally, would like to know joint distribution

$$F_{t_1,\ldots,t_n}(x_{t_1},\ldots,x_{t_n}) = P(X_{t_1} \le x_{t_1},\ldots,X_{t_n} \le x_{t_n}),\ \forall t_1,\ldots,t_n$$

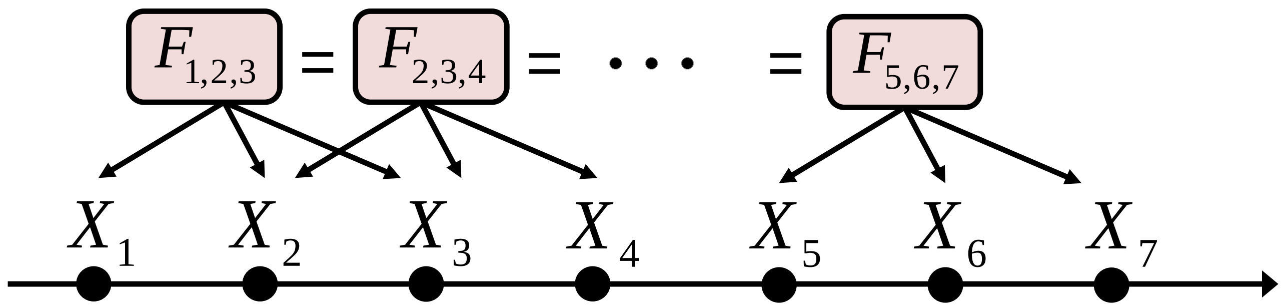
  for *any* subset of RVs
  - E.g. $F_{1,2,3}(x_1, x_2, x_3)$ describes how X_1, X_2, X_3 are related $\Longleftrightarrow$



- However, joint distr. is impossible to use unless we impose some *restrictions*

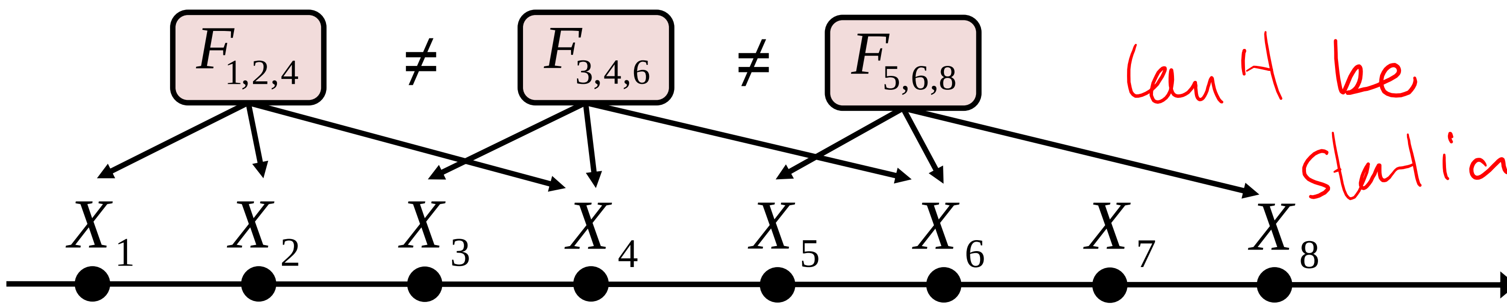
Strict Stationarity

- TS is called *strictly stationary* if its *joint distr. stays the same for any time-shift*
  - Formally: $F_{t_1,\ldots,t_n}(\cdot) = F_{t_1+s,\ldots,t_n+s}(\cdot), \forall\, s, t_1, \ldots, t_n$
  - Graphically:

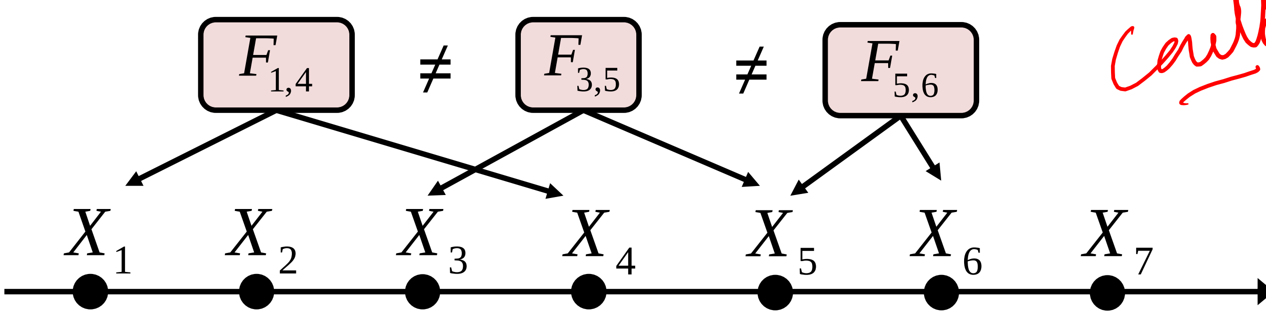


Example

- Can these TSs be strictly stationary?



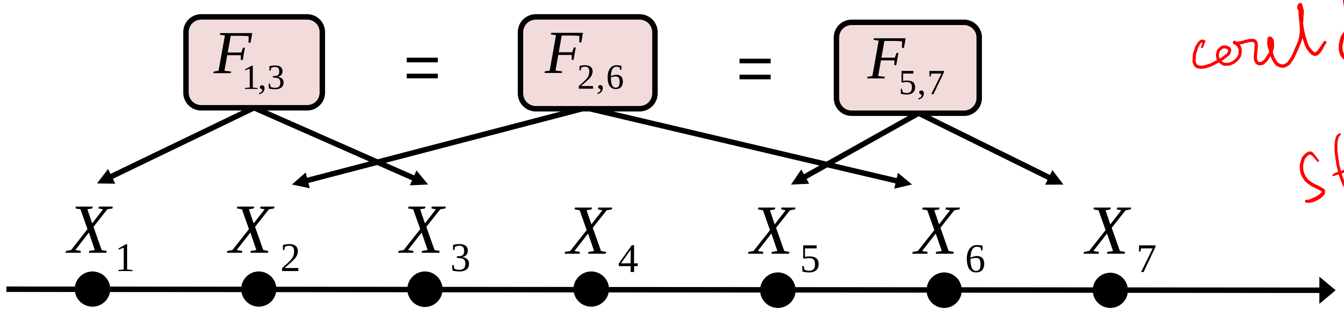
can't be stationary



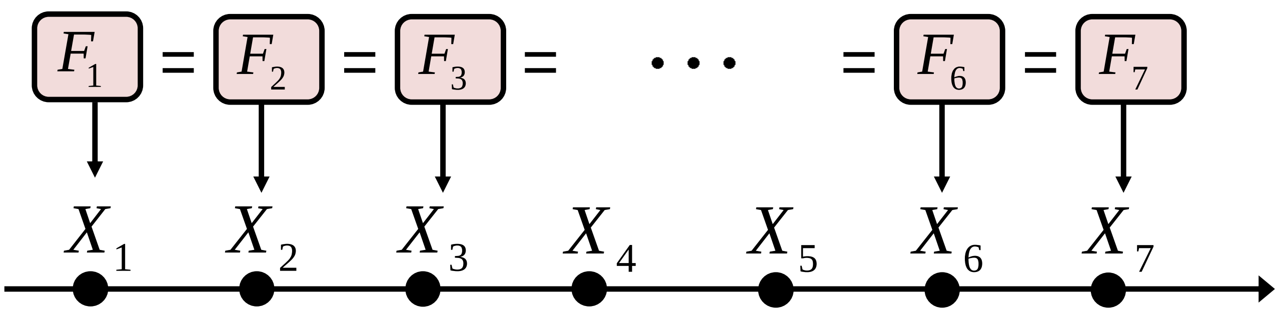
could be stationary

Example

- Can these TSs be strictly stationary?



could be stationary



could be stationary

Example

- Consider i.i.d. sequence $\{X_t\}$
  - Is this TS strictly stationary?

If $\{X_t\}$ is iid $\Rightarrow F_{t_1,\ldots,t_n}(\cdot\ \cdot\cdot) = F_{t_1}(\cdot)\times\cdots\times F_{t_n}(\,) =$

$= F(\cdot)\times\cdots\times F(\,)$

This holds for <u>any</u> set of indeces ($\forall$ collection of #n RV's)

$\Rightarrow F_{t_1,\ldots,t_n}(\cdot) = F_{t_1+s,\ldots,t_n+s}(\cdot) \Rightarrow$ TS is strictly stationary

Mean & Variance Functions

- TS *mean function* (μ_t) defined as mean of RV X_t at different times (t)

$$\mu_t = \mathrm{E}[X_t], \forall t$$

- *Variance function* (σ_t^2) defined similarly

$$\sigma_t^2 = \mathrm{V}[X_t], \forall t$$

  - For TS to be strictly stationary, mean & var functions *must be* constant w.r.t. time. Why?

by strict stationarity $\Rightarrow$ marginals are the same $\forall$ RV $\Rightarrow$ mean, var are the same

Covariances

- Even with strict stationarity, joint distr. is difficult to work with
  - Need to model $F_{X_1,\ldots,X_n}(\cdot)$ for *every* $n \geq 1$
- Simpler way of describing relationship between RVs is to look at their *covariance*
  - Covariance only measures *linear dependence*
    - $X \perp Y \rightarrow Cov(X,Y)=0$ but $Cov(X,Y)=0 \not\rightarrow X \perp Y$
    - Only if X,Y follow *multivariate Normal* distr. do we have $Cov(X,Y)=0 \Leftrightarrow X \perp Y$

Example

- Let $X \sim \mathrm{N}(0,1)$ and $Y = X^2$
  - Find $Cov(X,Y)$ $= \mathbb{E}[XY] - \cancel{\mathbb{E}[X]}^{0}\,\mathbb{E}[Y] =$
  - Is $X \perp Y$?

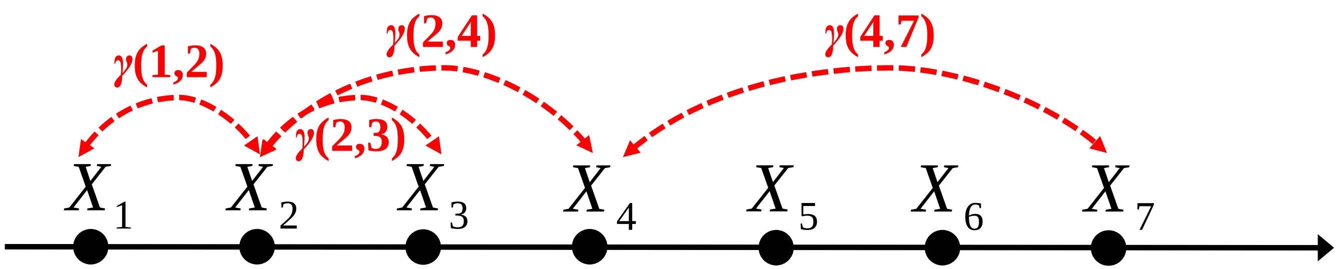
$= \mathbb{E}[X \cdot X^2] = \mathbb{E}[X^3] = 0$

No, because Y is a deterministic function of X

Autocovariance Function

- *Autocovariance function* (ACVF) of a TS measures covariance for *all pairs of RVs*
  - Defined as $\gamma(s,t) = Cov[X_s, X_t], \quad \forall s,t \in \mathbf{Z}$

$$\Leftrightarrow \gamma(s,t) = E[(X_s - \mu_s)(X_t - \mu_t)] = E[X_s X_t] - \mu_s \mu_t$$



  - Note that $Cov[X_t, X_t] = Var[X_t] = \sigma_t^2$

Autocovariance Function

- Assume we want to describe dependence of stationary TS $X_1,\ldots,X_n$
  - To do it *perfectly* we would need joint distr. $F_{X_1,\ldots,X_n}(\cdot)$ i.e. n-dimensional function
  - If we focus on *linear dependence* we only need $n \times n$ covariance matrix

$$
\begin{array}{c|cccc}
 & X_1 & X_2 & \cdots & X_n \\
\hline
X_1 & \gamma(1,1) & \gamma(1,2) & \cdots & \gamma(1,n) \\
X_2 & \gamma(2,1) & \gamma(2,2) & \cdots & \gamma(2,n) \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
X_n & \gamma(n,1) & \gamma(n,2) & \cdots & \gamma(n,n)
\end{array}
$$

Autocorrelation Function

- Covariance is difficult to interpret numerically $\rightarrow$ use correlation instead
- *Autocorrelation function* (ACF) of TS is

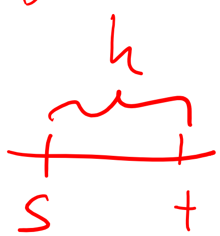
$$\rho(s,t) = \frac{\gamma(s,t)}{\sqrt{\gamma(s,s)\gamma(t,t)}} = \frac{Cov[X_s, X_t]}{\sqrt{Var[X_s]Var[X_t]}} = \frac{\gamma(s,t)}{\sqrt{\sigma_s^2\sigma_t^2}}$$

  - Always have $-1 \le \rho(s,t) \le +1$
- Essentially, ACF determines the *(linear) dependence structure* of a TS

Strict Stationarity

- For *strictly stationary* TS, ACVF *depends only on time difference* between RVs
  - I.e. $\gamma(s,t)=\gamma(h),\ \forall\{s,t:|s-t|=h\}$
  - Proof: without loss of generality, assume

$t = s+h$



$\gamma(s,t) = Cov(X_s, X_t)$

only depends on joint distr. $F_{s,t}$ of (X_s, X_t) which, by strict stationarity, is equal to joint $F_{0,h}$ of (X_0, X_h)

$= Cov(X_0, X_h) = \gamma(0,h) = \gamma(h)$

(same argument $\forall\ s-t=h$)

Weak Stationarity

- TS called *weakly stationary* (or just *stationary*) if:

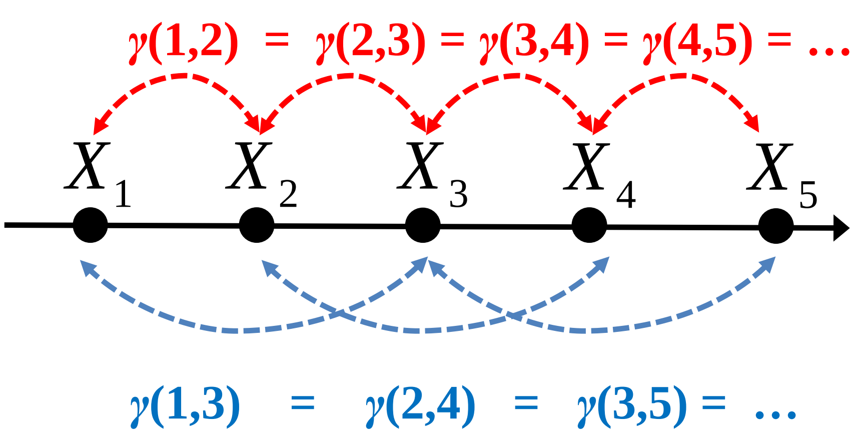
  (i) μ_t is constant (indep. of t)

  (ii) $\gamma(s,t) = \gamma(h)$ depends on s, t only through their difference $h = |s - t|$

- Strict stationarity $\rightarrow$ (weak) stationarity, but
- (Weak) Stationarity $\not\rightarrow$ Strict stationarity in general, ***unless*** TS is Normally distributed

Weak Stationarity

- For weakly stationary series, all auto-covariances/correlations over the same time-distance are equal:



	X_1	X_2	X_3	X_4	
X_1	$\gamma(1,1)$	$\gamma(1,2)$	$\gamma(1,3)$	$\gamma(1,4)$	$\cdots$
X_2	$\gamma(2,1)$	$\gamma(2,2)$	$\gamma(2,3)$	$\gamma(2,4)$	$\cdots$
X_3	$\gamma(3,1)$	$\gamma(3,2)$	$\gamma(3,3)$	$\gamma(3,4)$	$\cdots$
X_4	$\gamma(4,1)$	$\gamma(4,2)$	$\gamma(4,3)$	$\gamma(4,4)$	$\cdots$
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	

$\gamma(h=2)$ $\gamma(h=1)$ $\gamma(h=0)=\sigma^2$

Example (White Noise)

- Series of *uncorrelated* RVs with constant mean (μ) and variance (σ^2) is called *White Noise*, denoted by $W_t \sim \mathrm{WN}(\mu, \sigma^2)$
  - Find ACF of $\{W_t\}$

$\gamma(0) = \sigma^2$, & $\gamma(h) = 0 \quad \forall h \neq 0$

$\mathbb{E}[W_t] = \mu, \quad \forall t$

$$\Rightarrow \rho(h) = \frac{\gamma(h)}{\gamma(0)} = \begin{cases} 1, & \text{if } h = 0 \\ 0, & \text{if } h \neq 0 \end{cases}$$

Example (Random Walk)

- Consider series with $X_0 = 0$ and

$$X_t = X_{t-1} + W_t, \;\; \forall t \geq 1, \text{ for } W_t \sim WN(0, \sigma_w^2)$$

- Find mean & variance functions

$$\mu_t = \mathbb{E}[X_t] = \mathbb{E}[X_{t-1} + W_t] = \mathbb{E}[X_{t-1}] + \mathbb{E}[\cancel{W_t}]^{0} \Rightarrow$$

$$\Rightarrow \mu_t = \mu_{t-1} \;\&\; \mu_0 = \mathbb{E}[X_0] = 0 = \mu_1 = \mu_2 = \dots \Rightarrow \mu_t = 0, \forall t$$

$$\sigma_t^2 = \mathbb{V}[X_t] = \mathbb{V}[X_{t-1} + W_t] = \underbrace{\mathbb{V}[X_{t-1}]}_{\sigma_{t-1}^2} + \underbrace{\mathbb{V}[W_t]}_{\sigma^2} + 2\,\mathrm{Cov}[\cancel{X_{t-1}, W_t}]^{=0}$$

$$= \sigma_{t-1}^2 + \sigma_w^2 + 0 \Rightarrow \sigma_t^2 = (\sigma_{t-2}^2 + \sigma_w^2) + \sigma_w^2 = \dots = t \cdot \sigma_w^2$$

$$= \cancel{\sigma_0^2}^{0} + \overbrace{\sigma_w^2 + \dots + \sigma_w^2}^{t \text{ times}}$$

Example (cont'd) $X_t = \sum_{i=1}^{t} W_i + \cancel{X_0}^{=0} = X_s + \sum_{i=s+1}^{t} W_i$

- Find the autocovariance function

Let's assume $s<t$, so $\gamma(s,t) = Cov(X_s, X_t) =$

$$= Cov\left(X_s, X_s + \sum_{i=s+1}^{t} W_i\right) = \underbrace{Cov(X_s, X_s)}_{=Var(X_s) = s\cdot\sigma_w^2} + \sum_{i=s+1}^{t} \underbrace{Cov(X_s, W_i)}_{=0,\ \forall i \geq s+1}$$

$$\Rightarrow \gamma(s,t) = s\cdot\sigma_w^2, \quad \forall\ s<t$$

In general, $\gamma(s,t) = \min\{s,t\}\cdot\sigma_w^2, \quad \forall\ s,t$

& ACF is $\rho(s,t) = \dfrac{\gamma(s,t)}{\sqrt{\gamma(s,s)\gamma(t,t)}} = \dfrac{\min\{s,t\}\cdot\cancel{\sigma_w^2}}{\sqrt{s\cdot\cancel{\sigma_w^2}\cdot t\cdot\cancel{\sigma_w^2}}} = \dfrac{\min\{s,t\}}{\sqrt{s\cdot t}}$

Linear & Gaussian Processes

- TS $\{X_t\}$ defined as (possibly ∞) linear combination of WN called a *linear process*

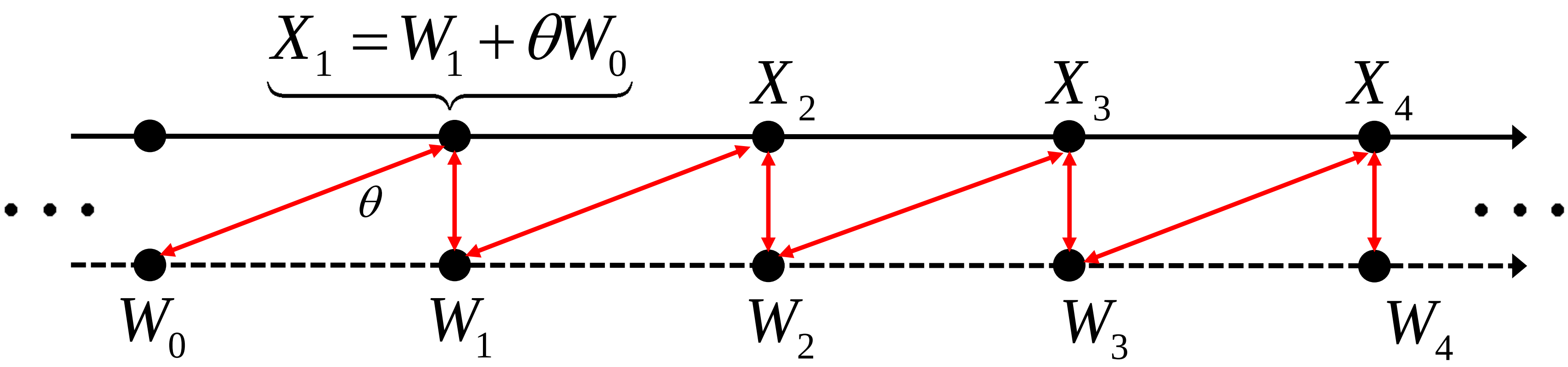
$$X_t = \mu + \sum_{j=-\infty}^{+\infty} \psi_j W_{t-j}, \quad \text{where} \quad \sum_{j=-\infty}^{+\infty} |\psi_j| < \infty$$

  - *Any* stationary TS can be expressed as a linear process
- If every collection of RVs $(X_{t_1}, \ldots, X_{t_n})$ has multivariate Normal distr. then TS is called a *Gaussian process*
  - If TS is Gaussian, its ACVF *completely defines* its dependence structure

Moving Average

- *Moving Average* (MA) TS defined as *linear combination of unobserved WN*

$$X_t = W_t + \theta W_{t-1}, \forall t \text{ where } \{W_t\} \sim \mathrm{WN}(0, \sigma_W^2)$$



Example (MA model)

- Consider $X_t = W_t + \theta W_{t-1},$ for $\{W_t\} \sim \mathrm{WN}(0, \sigma_w^2)$
  - Show $\{X_t\}$ is stationary

$\mu_t = \mathbb{E}[X_t] = \mathbb{E}[W_t + \theta \cdot W_{t-1}] = \mathbb{E}[\cancelto{0}{W_t}] + \theta\, \mathbb{E}[\cancelto{0}{W_{t-1}}] = 0, \ \forall t$

$\sigma_t^2 = \mathrm{Var}[X_t] = \mathrm{Var}[W_t + \theta W_{t-1}] = \underbrace{\mathrm{Var}[W_t]}_{=\sigma_w^2} + \theta^2 \cdot \overbrace{\mathrm{Var}[W_{t-1}]}^{\sigma_w^2} + 2\theta\, \mathrm{Cov}[\cancelto{0}{W_t, W_{t-1}}]$

$= (1+\theta^2) \cdot \sigma_w^2, \ \forall t$

let's assume $t = s+h = s+1 \quad (h=1) \Rightarrow$

$\Rightarrow \gamma(s,t) = \mathrm{Cov}(X_s, X_t) = \mathrm{Cov}(X_s, X_{s+1}) = \mathrm{Cov}(W_s + \theta W_{s-1}, W_{s+1} + \theta W_s)$

$= \mathrm{Cov}(\cancelto{0}{W_s, W_{s+1}}) + \theta\, \mathrm{Cov}(W_s, W_s) + \theta\, \mathrm{Cov}(\cancelto{0}{W_{s-1}, W_{s+1}}) + \theta^2 \mathrm{Cov}(\cancelto{0}{W_{s-1}, W_s})$

$= \theta \cdot \sigma_w^2, \ \forall s$

Examle (cont'd)

Let's assume $t = s+2 \Rightarrow \gamma(s,t) = Cov(X_s, X_{s+2}) =$

$$= Cov(W_s + \theta W_{s-1}, W_{s+2} + \theta W_{s+1}) =$$

$$= \cancel{Cov(W_s, W_{s+2})}^{0} + \theta \cancel{Cov(W_s, W_{s+1})}^{0} + \theta \cdot \cancel{Cov(W_{s-1}, W_{s+2})}^{0} + \theta^2 \cancel{Cov(W_{s-1}, W_{s+1})}^{0}$$

$$= 0 \quad , \forall s$$

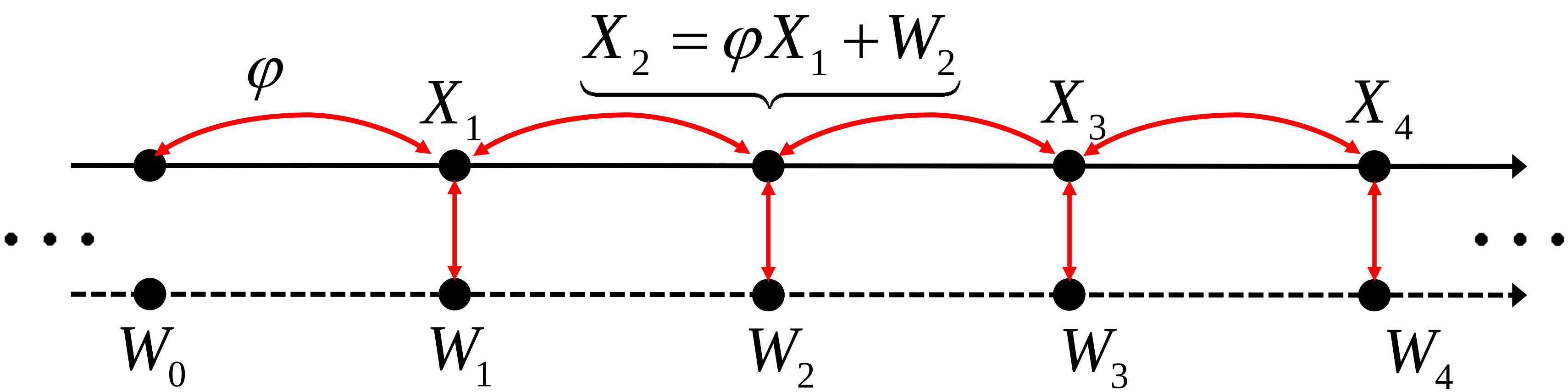
and similarly for $h \geq 3$

$$\Rightarrow \gamma(h) = \begin{cases} (1+\theta^2)\sigma_w^2 & , h=0 \\ \theta \sigma_w^2 & , h=1 \\ 0 & , h \geq 2 \end{cases} \Rightarrow \rho(h) = \begin{cases} 1, & h=0 \\ \frac{\theta}{1+\theta^2}, & h=1 \\ 0, & h \geq 2 \end{cases}$$

Autoregression

- *AutoRegressive* (AR) TS defined as *linear combination of its own past + noise*

$$X_t = \varphi X_{t-1} + W_t, \forall t \text{ where } \{W_t\} \sim \mathrm{WN}(0, \sigma_W^2)$$



Autoregression

- AR can be expressed as linear process
  - E.g. $X_t = \varphi X_{t-1} + W_t, \; \{W_t\} \sim \mathrm{WN}(0, \sigma_W^2)$

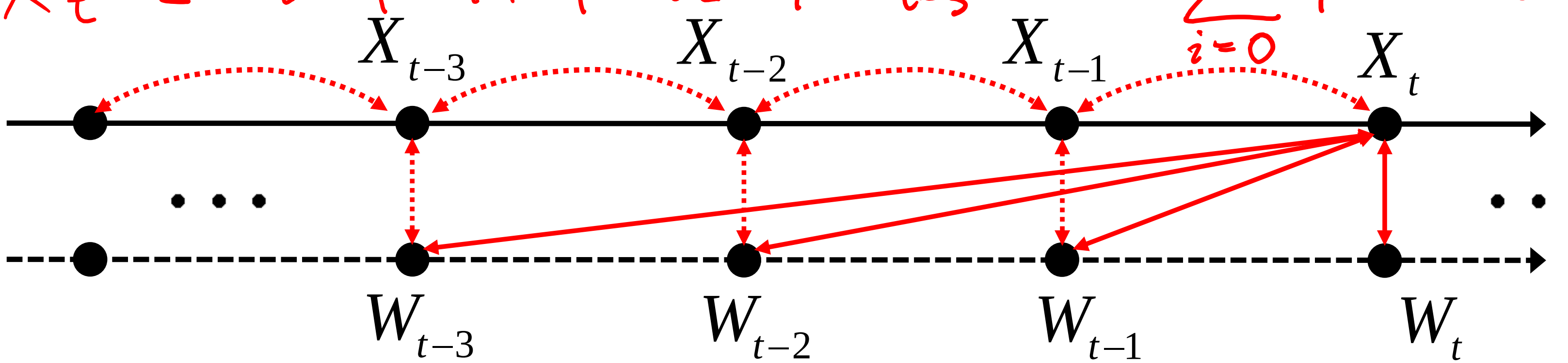
By iterated substitution: $X_t = \varphi \underline{X_{t-1}} + W_t =$

$$\Rightarrow X_t = \varphi(\varphi \cdot X_{t-2} + W_{t-1}) + W_t = \varphi^2 \cdot \underline{X_{t-2}} + \varphi W_{t-1} + W_t =$$

$$= \varphi^2 \cdot (\varphi X_{t-3} + W_{t-2}) + \varphi W_{t-1} + W_t = \varphi^3 X_{t-3} + \varphi^2 W_{t-2} + \varphi W_{t-1} + W_t$$

$$\vdots$$

$$X_t = W_t + \varphi W_{t-1} + \varphi^2 W_{t-2} + \varphi^3 W_{t-3} + \dots = \sum_{i=0}^{\infty} \varphi^i \cdot W_{t-i}$$



Example (AR model)

- Consider $X_t = .5X_{t-1} + W_t$, for $W_t \sim WN(0, \sigma_w^2)$
  - Show $\{X_t\}$ is stationary $\left\{ X_t = \sum_{i=0}^{\infty} \underbrace{(1/2)}_{=\varphi}^{i} \cdot W_{t-i} \right\}$

$$\mu_t = \mathbb{E}[X_t] = \mathbb{E}\left[\sum_{i=0}^{\infty} (1/2)^i W_{t-i}\right] = \sum_{i=0}^{\infty} (1/2)^i \cdot \mathbb{E}[\cancel{W}_{t-i}]^{=0} = 0$$

$$\sigma_t = Var[X_t] = Var\left[\sum_{i=0}^{\infty} (1/2)^i W_{t-i}\right] = \sum_{i=0}^{\infty} (1/2)^{2\cdot i} \underbrace{Var[W_{t-i}]}_{=\sigma_w^2} \quad \left(b/c\ Cov(W_i, W_j) = 0\ \forall i \neq j\right)$$

$$= \sigma_w^2 \cdot \sum_{i=0}^{\infty} (1/4)^i = \sigma_w^2 \cdot \frac{1}{1 - 1/4} = \sigma_w^2 \cdot 4/3$$

(geom. Series: $\sum_{i=0}^{\infty} a^i = \frac{1}{1-a}$, for $|a| < 1$)

Let's assume $t = s+1 \Rightarrow \gamma(s,t) = Cov(X_s, \underbrace{X_{s+1}}_{=\frac{1}{2}\cdot X_s + W_{s+1}}) =$

$$= Cov(X_s, \frac{1}{2}X_s + W_{s+1}) = \frac{1}{2}\underbrace{Cov(X_s, X_s)}_{Var[X_s]} + Cov(X_s, W_{s+1})$$

$$= \frac{1}{2}\underbrace{Var(X_s]}_{= 4/3 \cdot \sigma_w^2} + Cov\left(\sum_{i=0}^{\infty}(1/2)^i \cdot W_{s-i}, W_{s+1}\right)$$

$$= 2/3 \cdot \sigma_w^2 + \sum_{i=0}^{\infty} Cov(\cancel{W_{s-i}, W_{s+1}})^{0} \Rightarrow \forall s, \ \gamma(s,s+1) = \frac{2}{3}\sigma^2 = \gamma(1)$$

Let's assume $t = s+2 \Rightarrow \gamma(s,t) = Cov(X_s, \underline{X_{s+2}}) =$

$$= Cov(X_s, (\tfrac{1}{2})^2 X_s + \tfrac{1}{2}W_{s+1} + W_{s+2}) =$$

$$= (\tfrac{1}{2})^2 Cov(X_s, X_s) + \tfrac{1}{2}Cov(\cancel{X_s, W_{s+1}})^{0} + Cov(\cancel{X_s, W_{s+2}})^{0}$$

$$= \frac{1}{4}\cdot\frac{4}{3}\sigma_w^2 = \frac{1}{3}\sigma_w^2, \ \forall s$$

⋮

similarly, $\gamma(h) = \left(\frac{1}{2}\right)^h \cdot \frac{4}{3}\cdot\sigma_w^2$

Estimating Correlation

- ACF describes TS dependence structure
  - In practice *don't know* $\gamma(s,t)$ / $\rho(s,t)$ → need to *estimate* from data
- TS data $(x_1,\ldots, x_n)$ is sequence of *single realizations* from RVs $(X_1,\ldots, X_n)$
- But how can we estimate $\gamma(s,t)$ if we only have one data point x_t for each RV X_t?
  - That's where *stationarity* comes into play!!

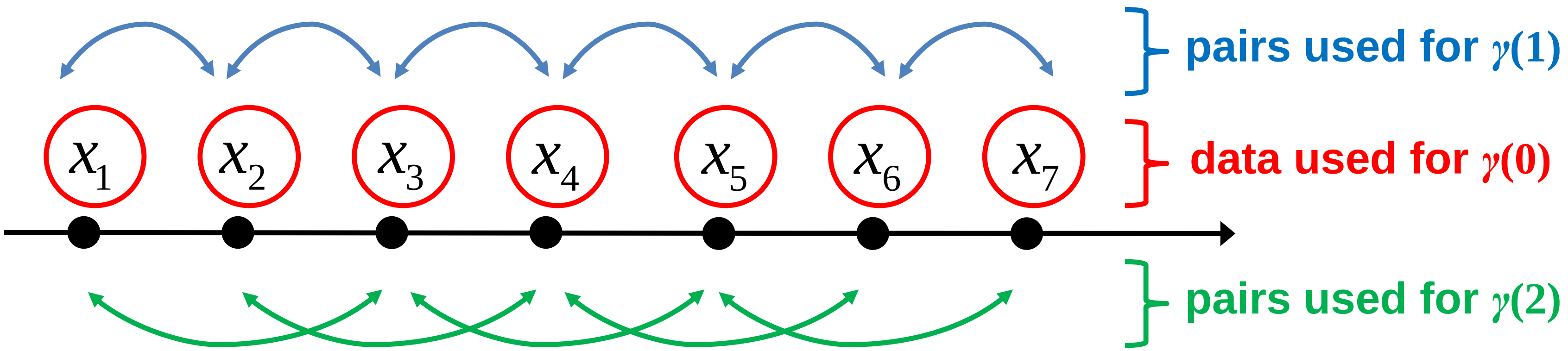
Estimating Correlation

- For *stationary* TS, $\gamma(s,t)$ depends on $h=|s-t|$ *alone* $\rightarrow$ $\gamma(s,t) = \gamma(t+h,t) = \gamma(h)$, for all t, $s=t+h$
    - E.g. $\gamma(1) = \mathrm{Cov}[X_2, X_1] = \mathrm{Cov}[X_3, X_2]$, etc…
- If TS is stationary $\rightarrow$ can use *all* data that are h times apart in order to estimate $\gamma(h)$
    - E.g. to estimate $\gamma(h) = \mathrm{Cov}[X_{t+h}, X_t]$ use (x_1, x_{1+h}), $(x_2, x_{2+h}),\ldots, (x_{n-h}, x_n)$, i.e. # $(n-h)$ pairs in total
    - Obviously, can only estimate $\gamma(h)$ up to lag $h=n-1$ (for now assume $\gamma(h)\approx 0$ for large h)

Estimating Correlation

- Estimate $\gamma(0)=\sigma^2$ (var) as usual, using *all* data
- Estimate $\gamma(1)$ using all pairs that are $h=1$ apart
- Estimate $\gamma(2)$ using all pairs that are $h=2$ apart

⋮



Estimating Correlation

For stationary TS $\{X_t\}$:

- Estimate μ by sample mean: $\bar{X} = \frac{1}{n}\sum_{t=1}^{n} X_t$
- Estimate $\gamma(h)$ by sample autocovariance:

$$\hat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-h}(X_{t+h} - \bar{X})(X_t - \bar{X}), \ \forall\ h = 0,1,\ldots,n-1$$

- Estimate $\rho(h)$ by sample ACF: $\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$

Estimating Correlation

- $\hat{\gamma}(h)$ is the sample estimate of the variance of the stationary TS $\sigma_t^2 = Var\left[X_t\right]$
- In formula for $\hat{\gamma}(0)$ *always* divide by n, despite having $(n-h)$ terms in sum
  - Neither n nor $(n-h)$ gives unbiased estimator,
  - Prefer dividing by n because it leads to non-negative definite $\hat{\gamma}(h)$
    - I.e. no linear combination $a_1X_1+\ldots+a_nX_n$ will have negative estimate of its variance

ACF Plot

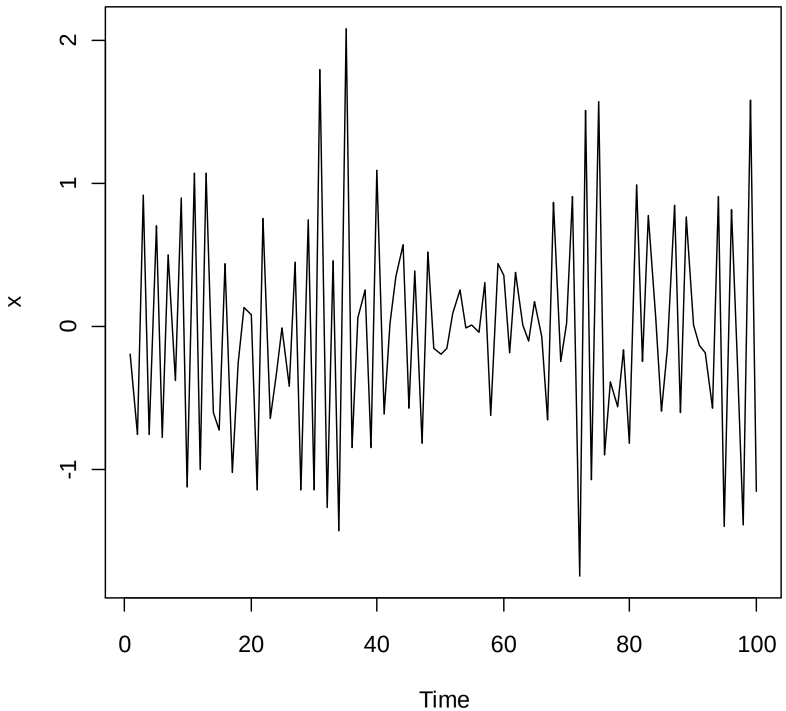
- ACF describes dependence structure of TS
- In practice, estimate $\gamma(h)$ from data

$$\hat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-h}(X_{t+h} - \bar{X})(X_t - \bar{X}), \text{ for } h = 0,1,\ldots,n-1$$

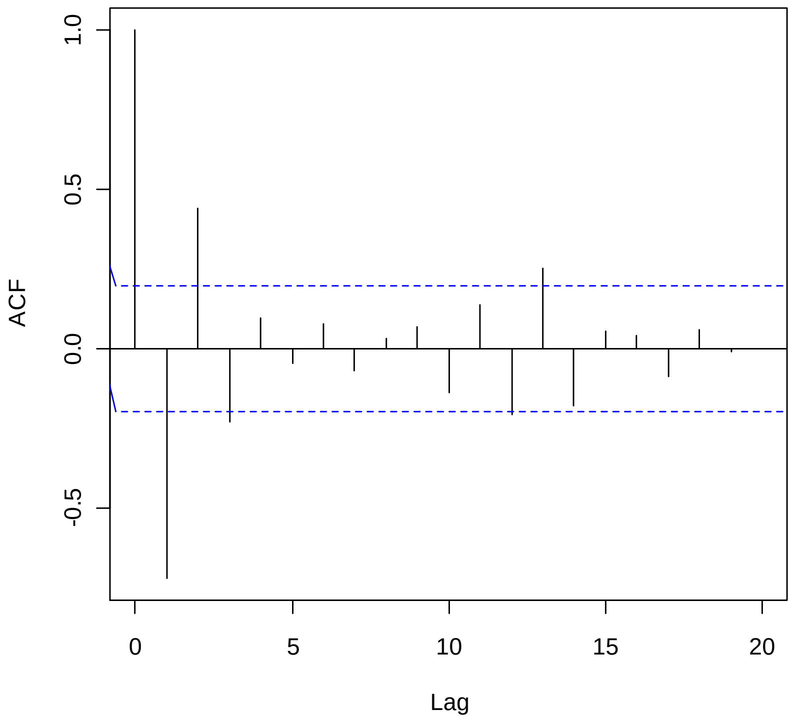
  - *Sample ACF* given by: $\hat{\rho}(h) = \dfrac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$
- Typically present results using *ACF plot*
  - Essentially a barplot of $\hat{\rho}(h)$

ACF Plot

- Time Series Data
  (moving average)



- Sample ACF Plot
  (`acf` function in R)



Estimating Correlation

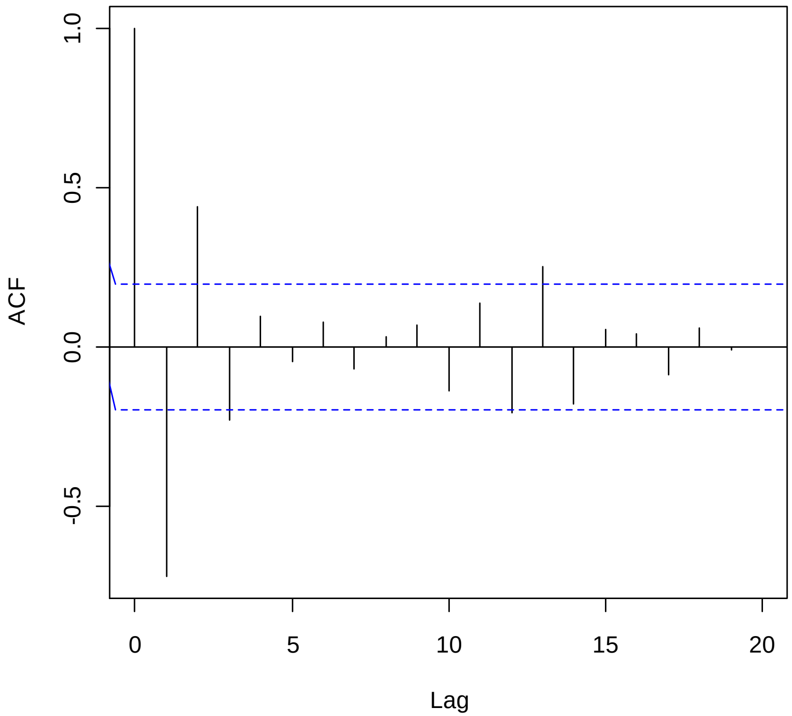
- If $\{X_t\}$ is White Noise (WN), then as $n \to \infty$

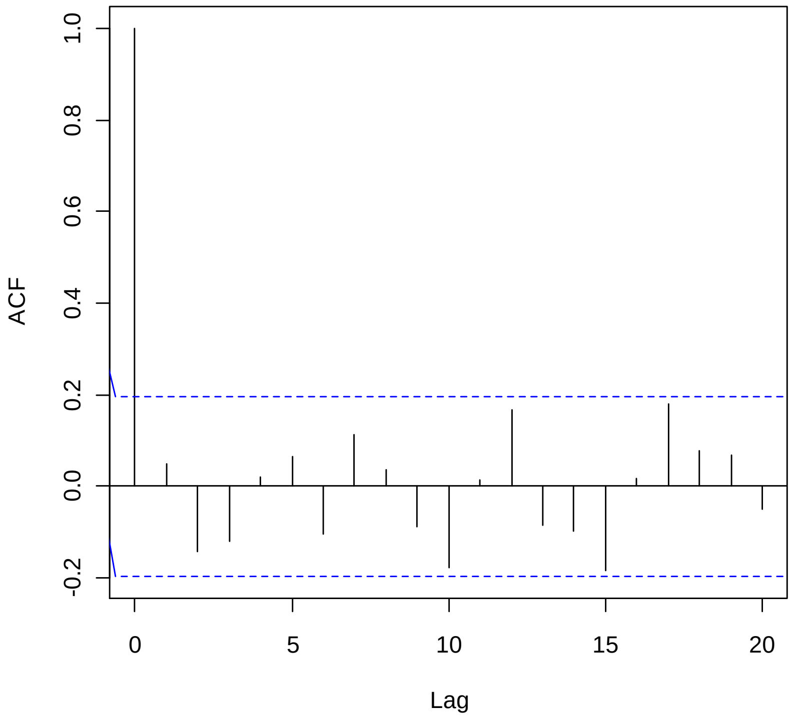
$$\hat{\rho}(h) \sim N\left(0, \frac{1}{n}\right), \forall h = 1, \ldots, H \text{ for a fixed } H$$

- Use result to check if TS is *uncorrelated*
  - Plot sample ACF for a number of lags
  - If TS is WN, then expect ~95% of autocorrelations to fall *within* 0±2×SD, i.e. within $\pm 2 / \sqrt{n}$
  - These WN "confidence bounds" are denoted by dashed lines in ACF plot

Example

- Sample ACF for MA
- Sample ACF for WN



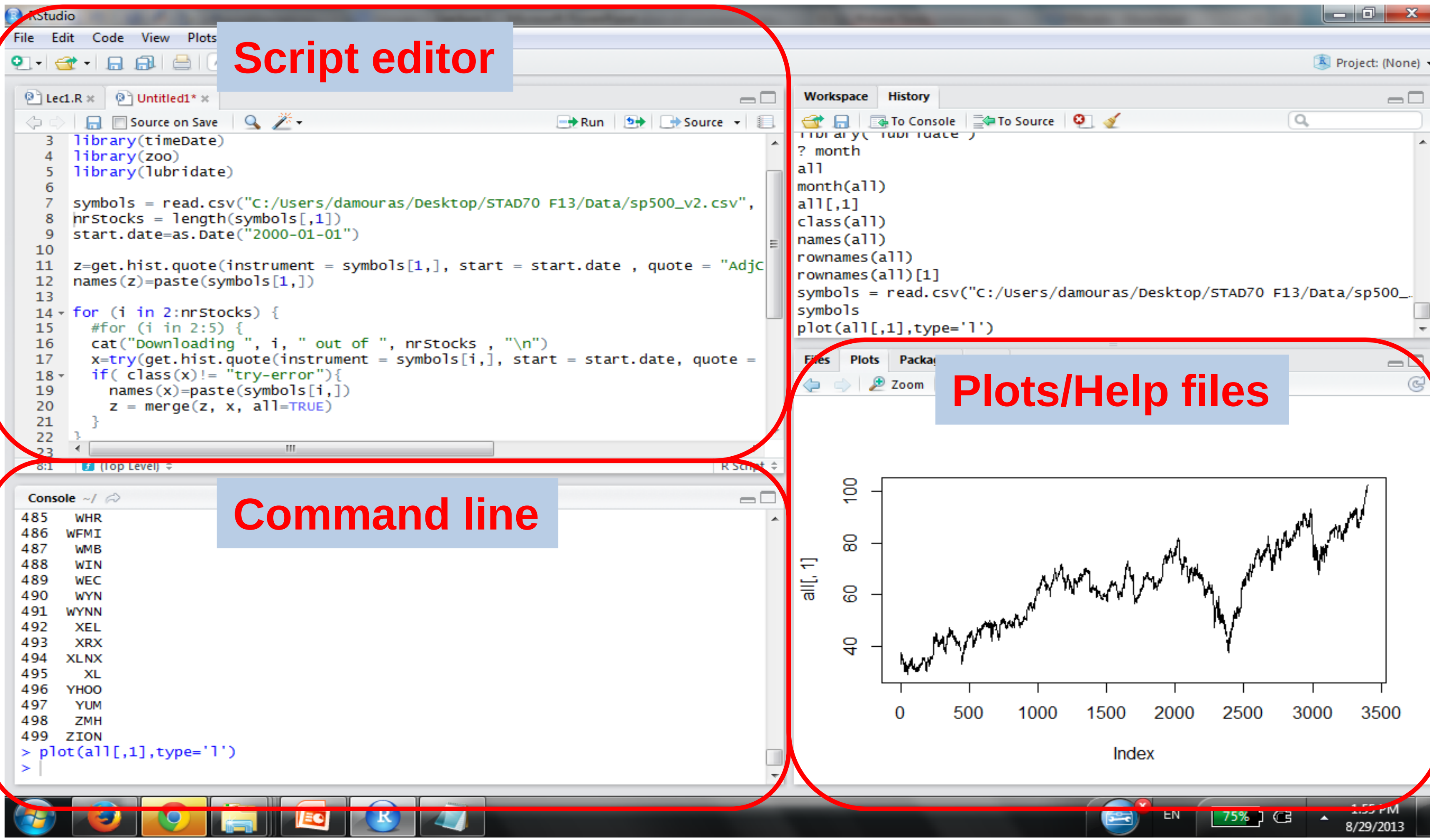


$n = 100$ for both series $\Rightarrow$ WN bounds are $\pm .2$

Statistical Programming

- We are going to use R for all practical applications
  - Download from http://cran.r-project.org/
- Work through R-Studio
  - Can edit, run, see results of R code all in one screen
  - Download from http://www.rstudio.com/ide/download/
- Both available for PC, Mac & Linux

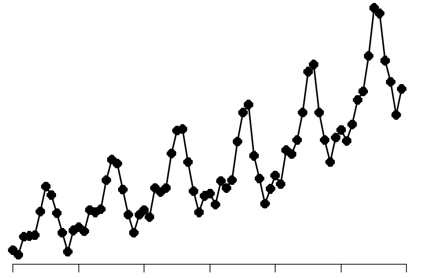
R Studio



Getting Started

- R-Studio Tutorials: https://www.rstudio.com/resources/training/online-learning/#R

- Data & code for all examples in textbook webpage http://www.stat.pitt.edu/stoffer/tsa4/
  - Load package “astsa” onto R to get data
  - Get sample code from http://www.stat.pitt.edu/stoffer/tsa4/Rexamples.htm



2. Exploratory Time Series Analysis

STAD57 F19

Sotirios Damouras

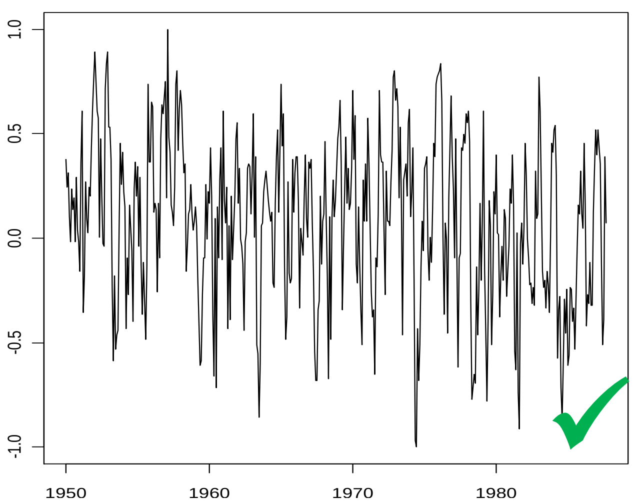
Exploratory TS Analysis

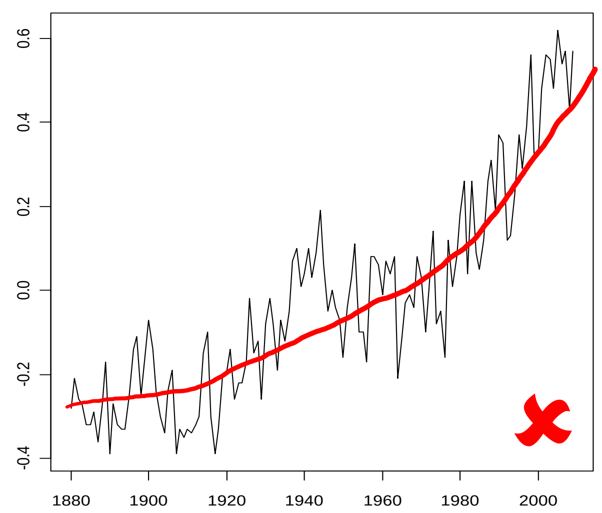
- Given TS data, want to understand their characteristics & prepare them for further analysis

- In particular, check if series is stationary; if not try to make it stationary
  - Check 3 things:
    - μ_t is constant
    - σ_t^2 is constant
    - $\gamma(s,t)$ is function of $h=s-t$

Exploratory Data Analysis

- Start with TS plot & check for constant mean (want series to be "flat")



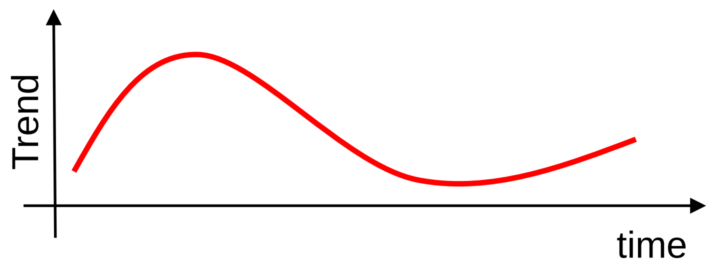


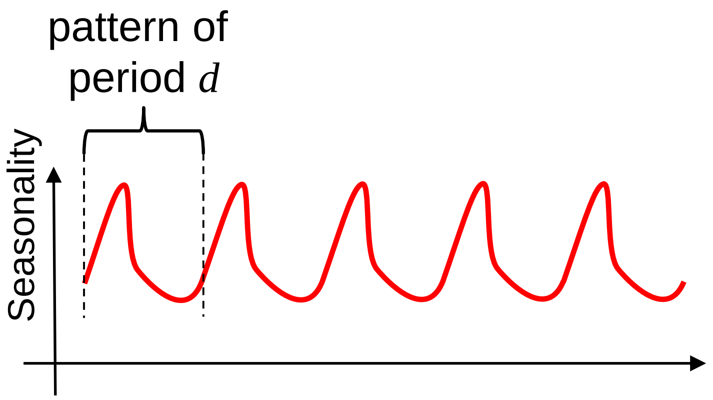
Trend & Seasonality

Mean function can fluctuate in two ways

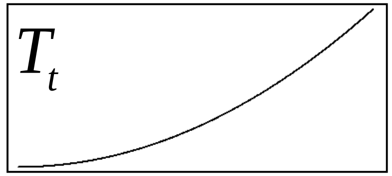
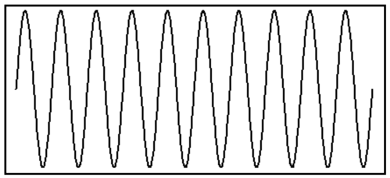
- *Trend* (T_t) gradual, non-repeating change in mean level
- *Seasonality* (S_t) periodic change in mean level
  - *Period* (d) cycle length

$$S_{t+d} = S_t, \ \forall t$$

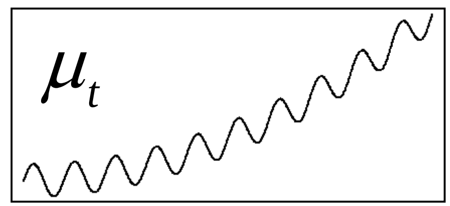




Modeling Trend & Seasonality

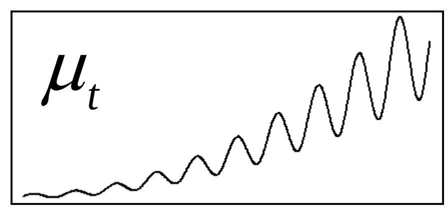
- For both trend  and seasonality S_t , two ways to *decompose* TS
  - *Additive:*

$$Y_t = \underbrace{T_t + S_t}_{\mu_t} + X_t \quad \Rightarrow$$



  - *Multiplicative*

$$Y_t = T_t \times S_t \times X_t \quad \Rightarrow$$



original series

trend component

seasonal component

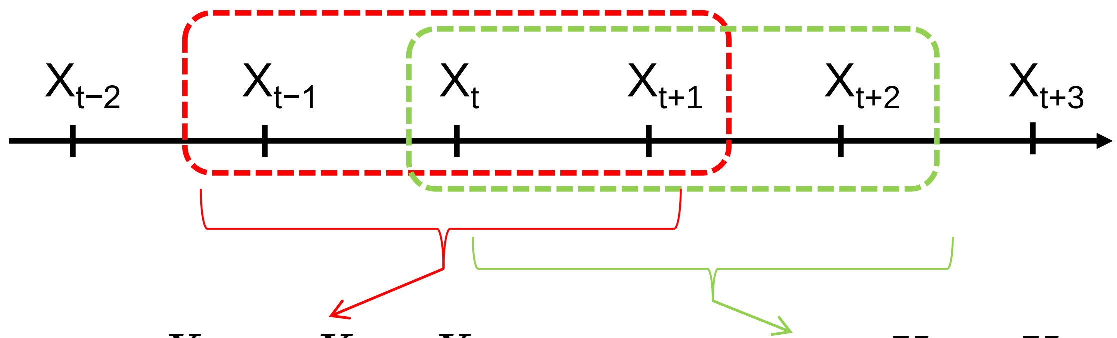
stationary remainder

Moving Average Smoother

- MA smoother suppresses TS variability and allows detection of trend & seasonality
  - For seasonality, window equal to period of cyclic pattern (or multiples thereof) gives smoothest MA, i.e. it will completely suppress pattern
    - Can use this to verify period of seasonal component
- If we detect either trend or seasonality, can try to remove them by:
  - Estimating & subtracting μ_t from series, or
  - Differencing series

Moving Average Smoother

- Average over sliding window of size q called *q-point moving average* (like a "local" mean)
  - E.g. 3-point MA $m_t = \frac{1}{3}\sum_{j=-1}^{1} X_{t-j}$, for $t = 2,..,n-1$



$$m_t = \frac{X_{t-1} + X_t + X_{t+1}}{3} \qquad m_{t+1} = \frac{X_t + X_{t+1} + X_{t+2}}{3}$$

Moving Average Smoother

- q-point MA for *odd* q defined as:

$$m_t = \frac{X_{t-k} + \ldots + X_t + \ldots + X_{t+k}}{q}, \text{ for } k = \frac{q-1}{2}$$

- q-point MA for *even* q defined as:

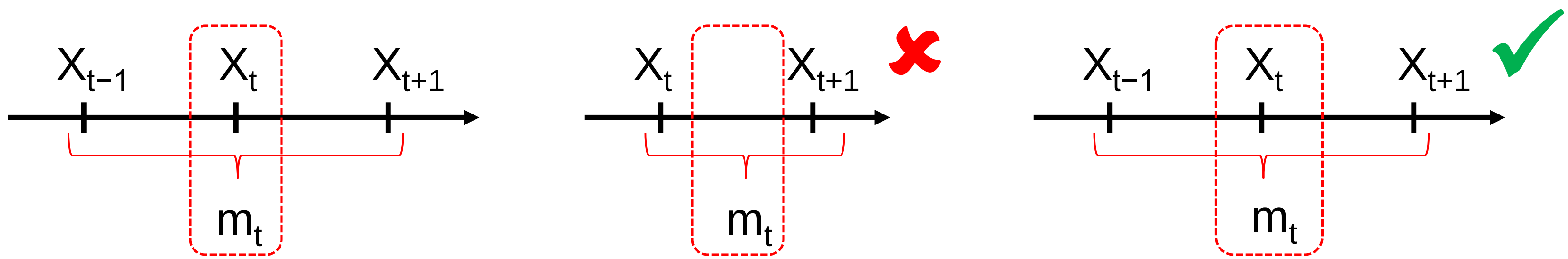
$$m_t = \frac{.5X_{t-k} + X_{t-k+1} + \ldots + X_t + \ldots + X_{t+k-1} + .5X_{t+k}}{q}, \text{ for } k = \frac{q}{2}$$

  - E.g. 4-point MA is

$$m_t = \frac{.5X_{t-2} + X_{t-1} + X_t + X_{t+1} + .5X_{t+2}}{4}$$

Moving Average Smoother

- Different method for odd/even q because we want MA to be *aligned* with time t
  - Odd (3-point)
  - Even (2-point)



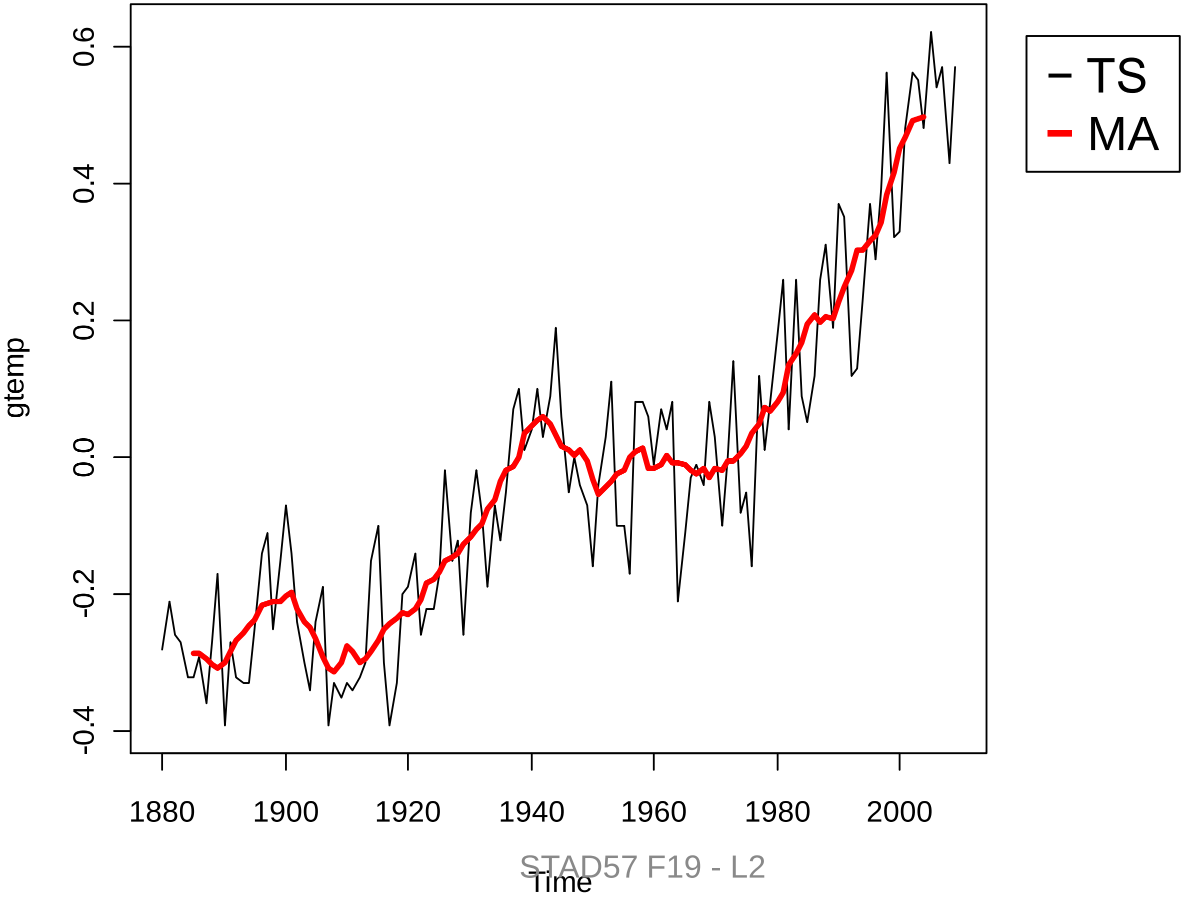
$= (X_t+X_{t+1})/2$

$= (.5X_{t-1}+X_t+.5X_{t+1})/2$

- Note that q-point MA not defined for first & last k observations (for either k=(q-1)/2 or q/2)
  - E.g. for $(X_1,\ldots,X_n)$, 5-point MA runs from m_3 to m_{n-2}

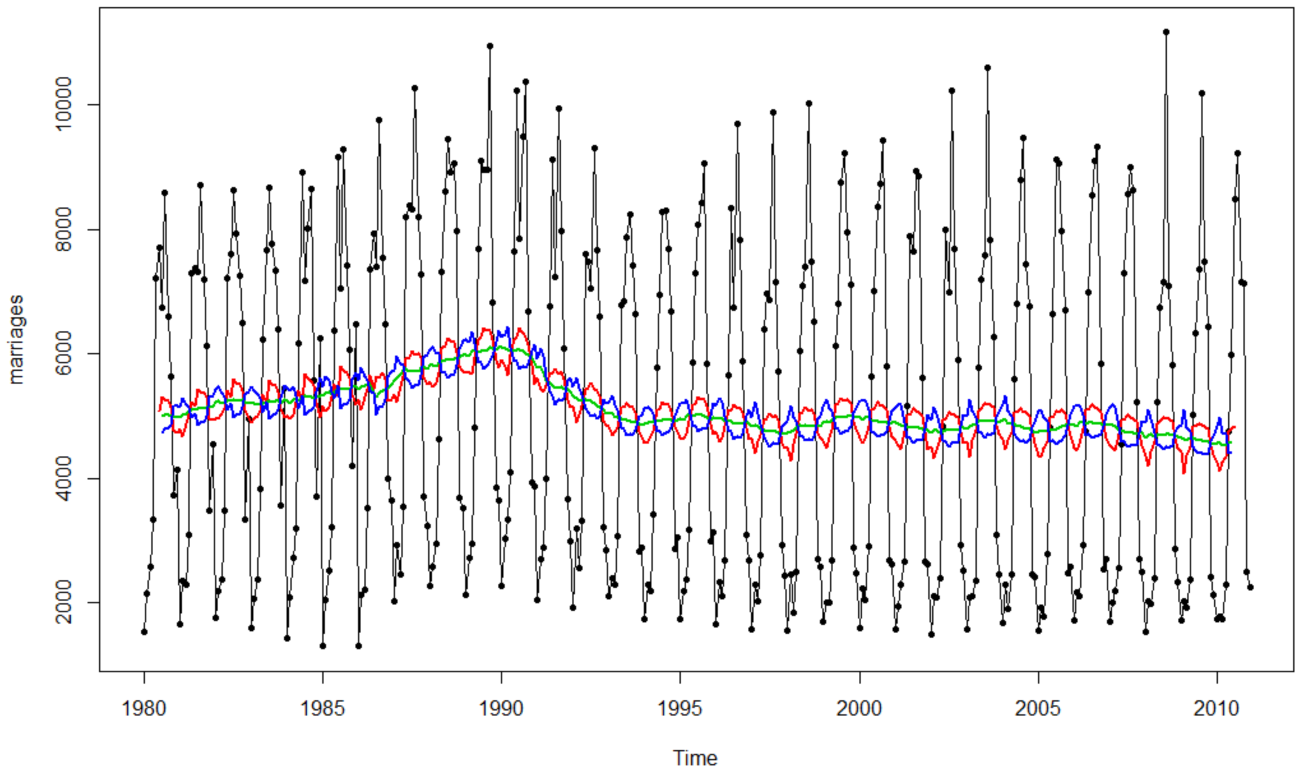
Example

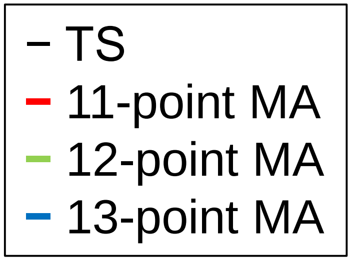
- Global mean temperature & 11-point MA



Example

- Ontario Marriages





TS seems to have annual cycle → seasonality period is 12

Filtering in R

- R function `filter()` allows us to calculate MA smoother and simulate MA & AR series
- A filter is a linear combination of a series with linear coefficients a_j
  - Filtered series given by: $y_t = \sum_{j=-k}^{k} a_j x_{t+j}$
- For q-point MA smoother, filter coeff's are:
  - For odd q: $a_j = 1/q, \ \forall j = -k, \ldots, k$ where $k = (q-1)/2$
  - For even q: $\begin{cases} a_{-k} = a_k = .5/q, \\ a_j = 1/q, \ \forall j = -k+1, \ldots, k-1 \end{cases}$ where $k = q/2$

Filtering in R

TS vector filter coeff. vector

- Use function in R as: `filter( X , a )`
  - E.g. 2-point MA: `filter(X, c(.5,1,.5)/2)`
- Can also use *one-sided* MA: $y_t = \sum_{j=0}^{k} a_j x_{t-j}$
  - Useful for simulating MA model data
  - E.g. Consider MA model $X_t = W_t + 2W_{t-1} + 3W_{t-2}$
    - Simulate series based on WN sequence $\{W_t\}$

```
X=filter( W, 1:3, side=1)
```

(Note: define coeff's in order a_0, a_1,…)

Filtering in R

- Use *recursive* filter for simulating AR series:

$$y_t = x_t + \sum_{j=1}^{k} a_j y_{t-j} = x_t + a_1 y_{t-1} + \ldots + a_k y_{t-k}$$

  - Recursive filter uses previous output values
  - E.g. Consider AR model $X_t = .5X_{t-1} - .3X_{t-2} + W_t$
    - Simulate series based on WN sequence $\{W_t\}$

```
X=filter( W, c(.5,-.3), method='recursive')
```

(Note: define coeff's in order a_1, a_2,…)

Detrending

- Plot TS and/or its MA: if μ_t seems changing, estimate μ_t and subtract it from series

$$\underbrace{Y_t}_{\text{observed TS}} = \underbrace{\mu_t}_{\text{trend}} + \underbrace{X_t}_{\text{stationary 0-mean TS}} \Rightarrow \boxed{\text{estimate trend } \hat{\mu}_t} \Rightarrow \text{work with } \underbrace{\hat{X}_t}_{\text{detrended TS}} = Y_t - \hat{\mu}_t$$

  - Since $E[Y_t] = \mu_t$, we have $E[\hat{X}_t] = E[Y_t - \hat{\mu}_t] \approx 0$
- How do we estimate μ_t?
  - For *deterministic* trend (i.e. function of t), can use regression

Estimating Trend

- Assume deterministic trend $\mu_t = f(t)$
  - For *linear* trend, use simple linear regression

$$Y_t = \beta_0 + \beta_1 t + X_t$$

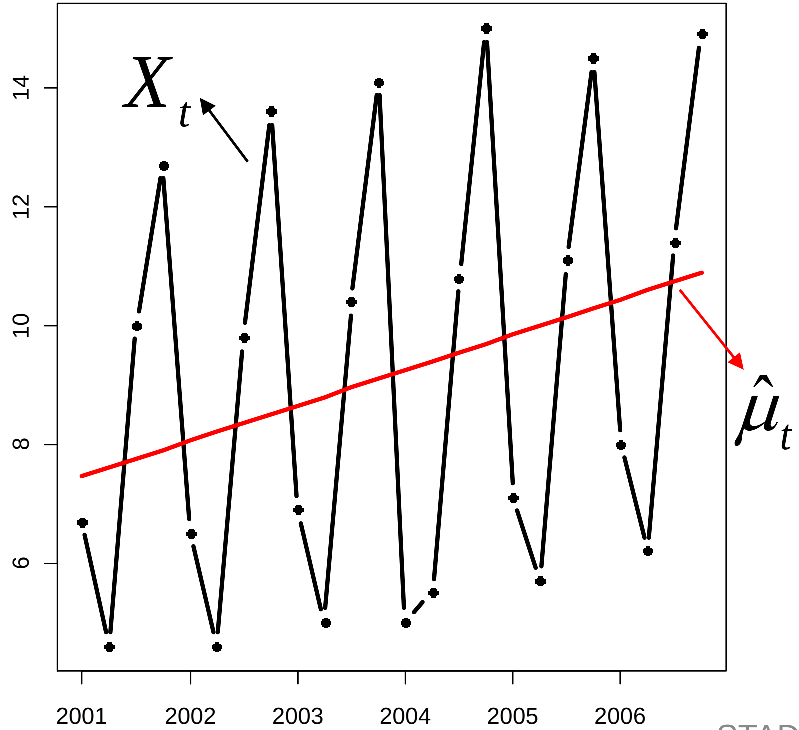
  - For nonlinear trend, can use polynomial

$$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \ldots + X_t$$

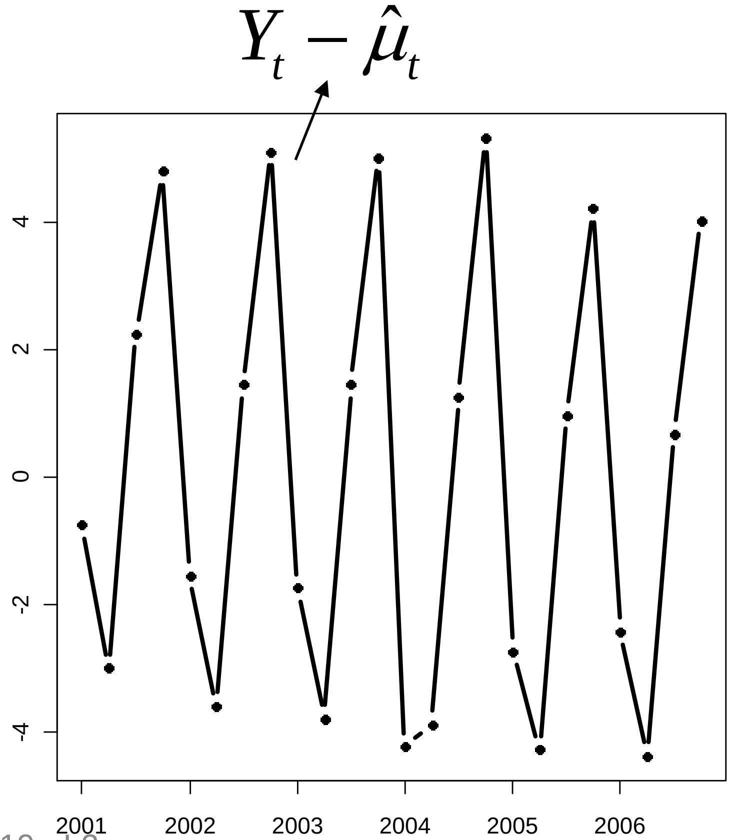
    or other nonlinear regression method
- Regression's explanatory variable is time (t), or some function thereof

Example

- Series w/ fitted linear trend

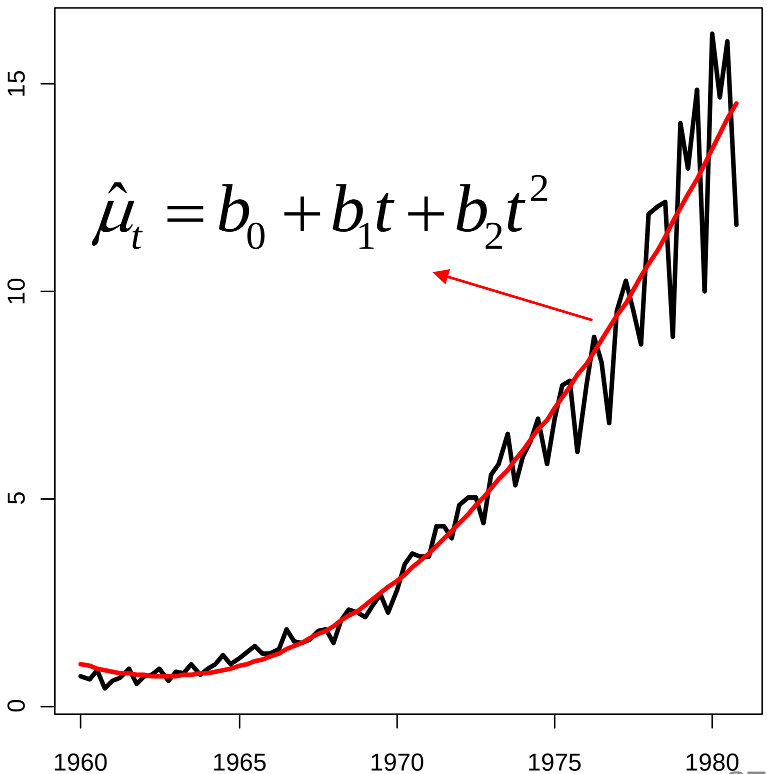


- Detrended TS

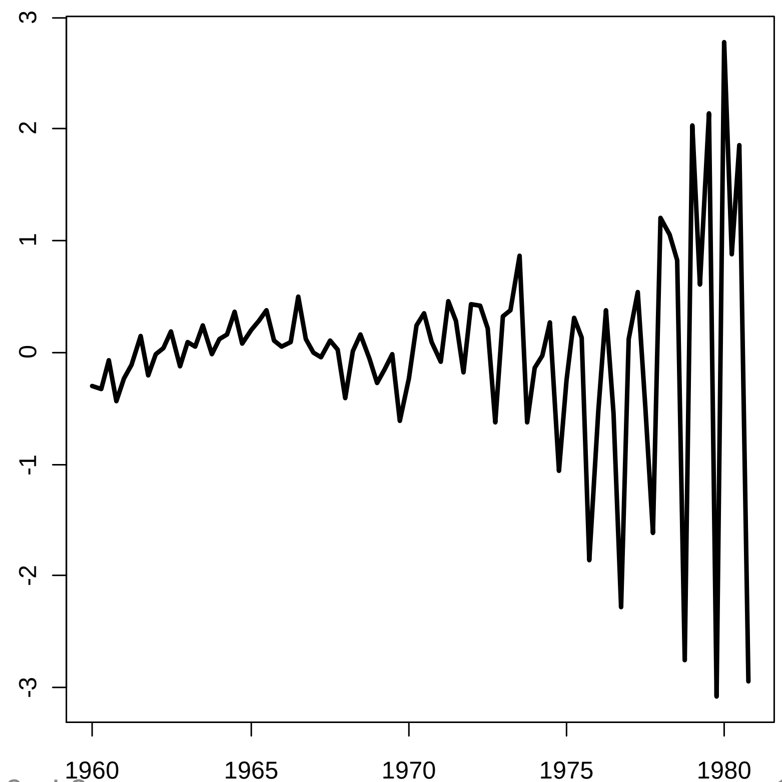


Example

- J&J sales with fitted quadratic trend



- De-trended TS



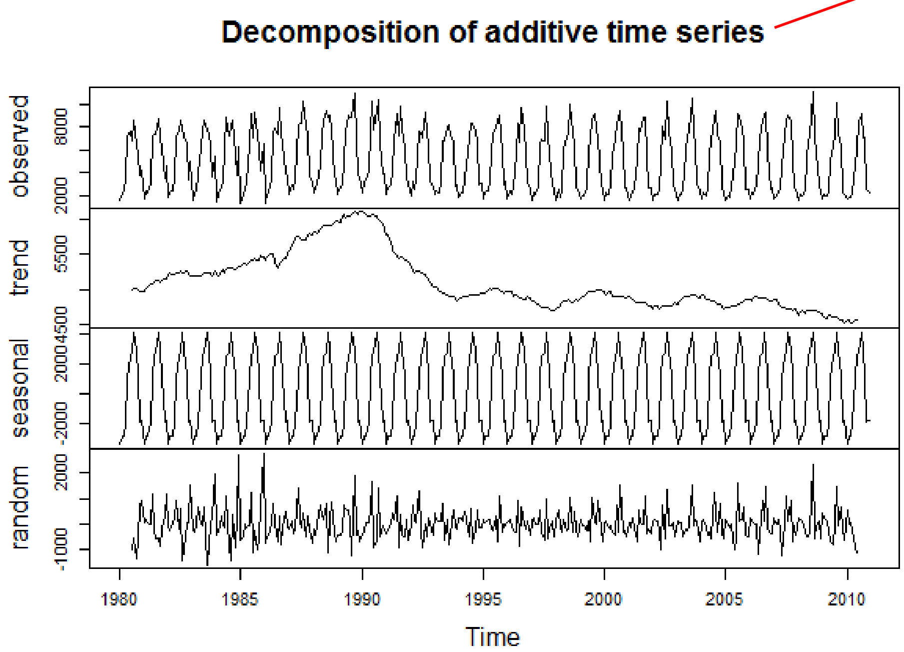
Regression in R

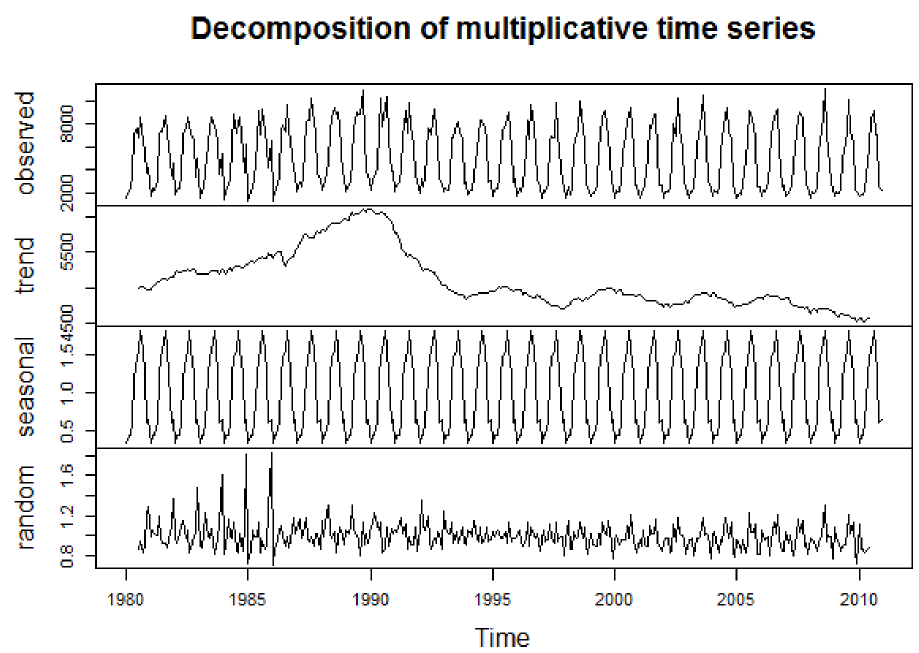
- Fit linear model $Y_i = \beta_0 + \beta_1 X_{1,i} + \ldots + \beta_p X_{p,i} + \varepsilon_i$ with command: $\underbrace{\texttt{lm}}_{\text{linear model fn}}(\ \underbrace{\texttt{Y}}_{\text{response variable}}\ \sim\ \underbrace{\texttt{X1 + ... + Xn}}_{\text{explanatory variables}}\ )$
  - Can save output to R object: `out=lm(Y~X1+…)`
  - Use `summary(out)` to see model summary
    - Coefficient & error std estimates, significance tests
  - Feed output to other functions to get:
    - ANOVA table: `aov(out)`
    - Fitted values: `fitted(out)`
    - Akaike's criterion: `AIC(out)`, etc

Example (TS decomposition)

- R function `decompose` (MA smoother for T & ANOVA for S)
  - E.g. marriages data

source	*% SS*
Trend	2.17
Seasonality	91.69
Random	5.48





Differencing

- Alternative to fitting linear trend is *differencing*:

$$Y_t = \beta_0 + \beta_1 t + X_t \quad \Rightarrow \quad (Y_t - Y_{t-1}) = \beta_1 + (X_t - X_{t-1})$$

  - If $\{X_t\}$ stationary $\rightarrow$ $\{X_t - X_{t-1}\}$ also stationary
- Differencing more appropriate when $\{X_t\}$ is random walk $\rightarrow$ $\{Y_t\}$ is *random walk* w/ drift
  - Use when $\{Y_t\}$ seems to “hover” around linear trend. In this case, detrending is not enough to make series stationary
- With differencing we end up with $\{X_t - X_{t-1}\}$

Differencing

- Define *backshift* operator *B* as $BX_t = X_{t-1}$ and extend it to powers so that $B^k X_t = X_{t-k}$
- Define 1st order *difference* operator ∇ as

$$\nabla X_t = (1-B)X_t = X_t - X_{t-1}$$

  - Extend it to higher (d th) order differences

$$\nabla^d X_t = (1-B)^d X_t$$

  by algebraically expanding operator $(1-B)^d$

  - E.g. $\nabla^2 X_t = (1-B)^2 X_t = (1-2B+B^2)X_t =$

$$= X_t - 2BX_t + B^2 X_t = X_t - 2X_{t-1} + X_{t-2}$$

Example (Random Walk with Drift)

$= \delta + (\delta + X_{t-2} + W_{t-1}) + W_t = 2\delta + X_{t-2} + W_{t-1} + W_t =$

$= 3\delta + X_{t-3} + W_{t-2} + W_{t-1} + W_t$

$\vdots$

$= t\cdot\delta + X_0 + \sum_{i=0}^{t-1} W_{t-i}$

- Consider series with X_0=0 and

  $X_t = \delta + X_{t-1} + W_t, \ \forall t \geq 1$, for $W_t \sim WN(0, \sigma_w^2)$

  - What are mean & variance functions of X_t

$E[X_t] = E\left[t\cdot\delta + \cancel{X_0}^{0} + \sum_{i=0}^{t-1} W_{t-i}\right] = t\cdot\delta + \sum E[\cancel{W}]^{0} \Rightarrow \mu_t = t\cdot\delta$, $\sigma_t^2 = V[X_t] =$

$= \ldots = t\cdot\sigma_w^2$

  - Show differenced series $\nabla X_t = X_t - X_{t-1}$ is stationary

$\nabla X_t = X_t - X_{t-1} = \delta + \cancel{X_{t-1}} + W_t - \cancel{X_{t-1}} = \delta + W_t$

where W_t is WN $\Rightarrow$ stationary $\Rightarrow$ $\delta + W_t$ also stationary

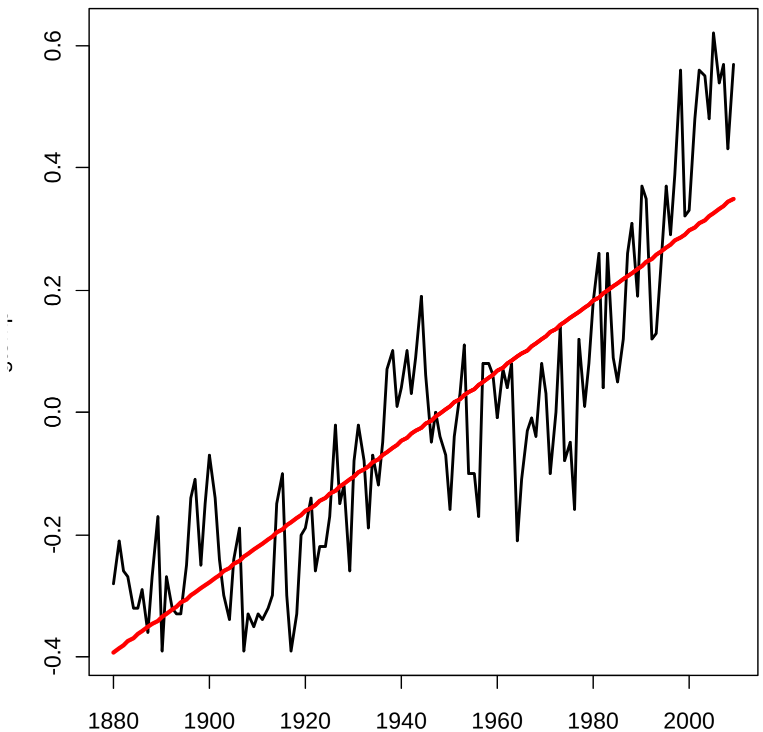
Differencing in R

- For 1st order differences, i.e. $\nabla X_t = X_t - X_{t-1}$ use `diff(X)`
- For dth order differences, i.e. $\nabla^d X_t = (1-B)^d X_t$ use `diff(X, differences=d)`
- For simple (1st order) differences *at lag k*, i.e. $(1-B^k)X_t = X_t - X_{t-k}$, use `diff(X,lag=k)`

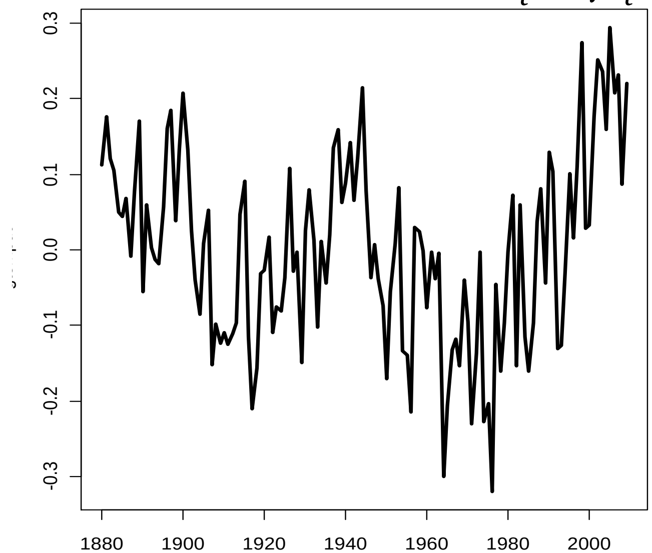
Example

- Global Temperature

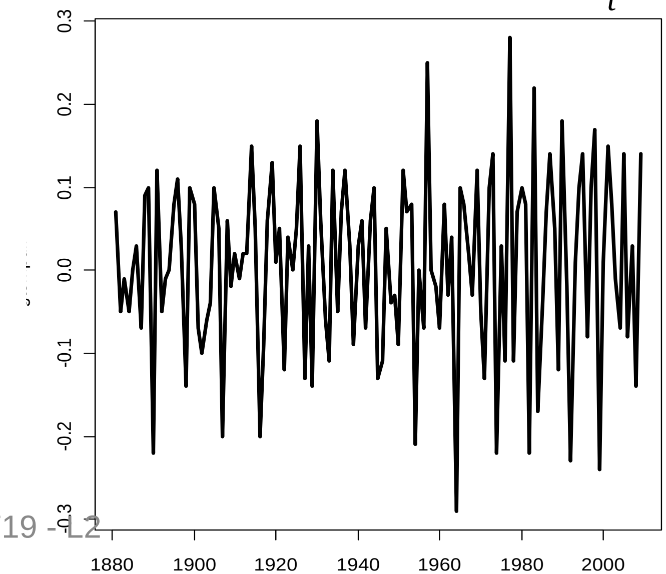
TS w/ fitted linear trend



de-trended series $Y_t - \hat{\mu}_t$

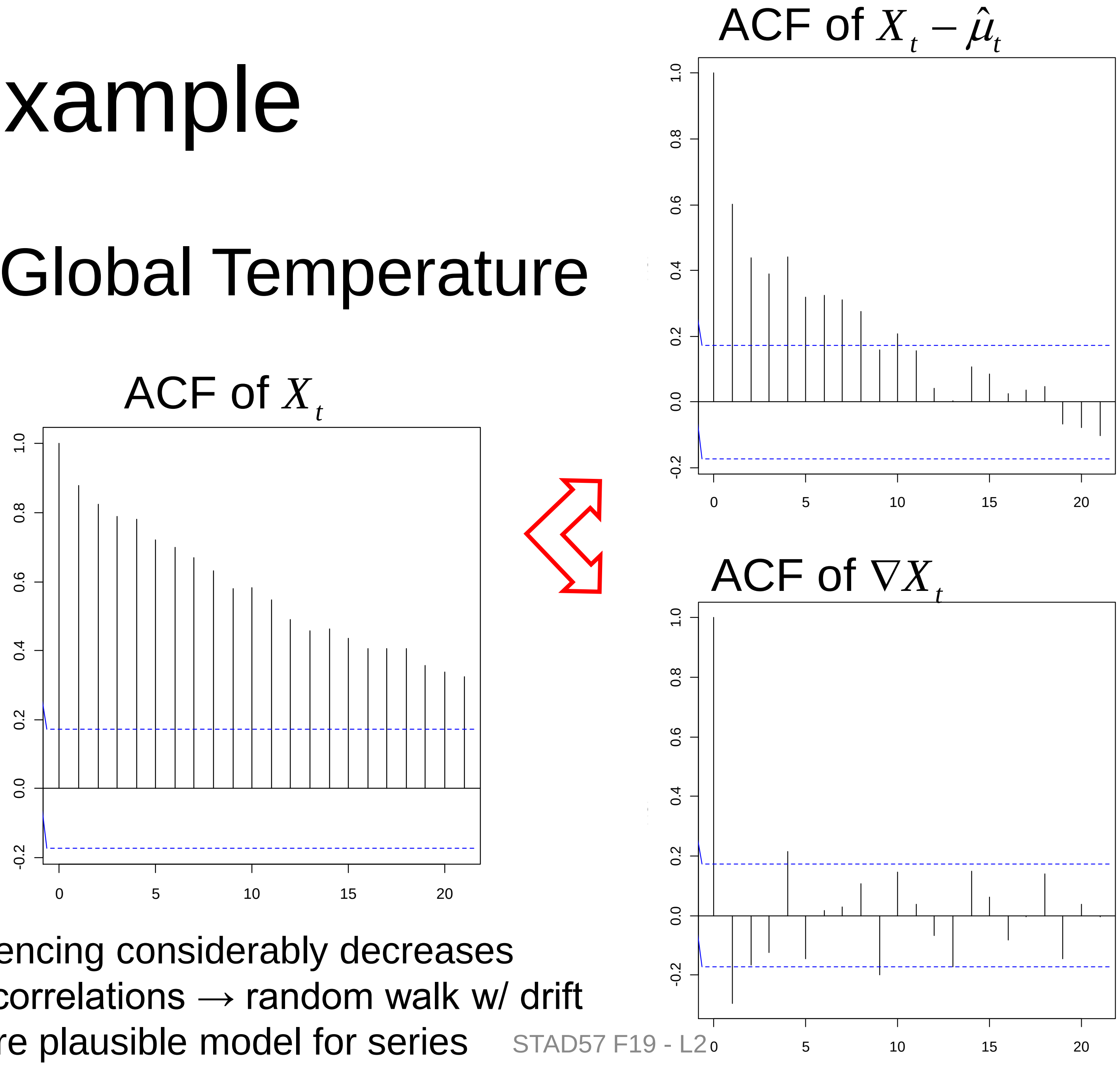


differenced series ∇Y_t



Example

- Global Temperature

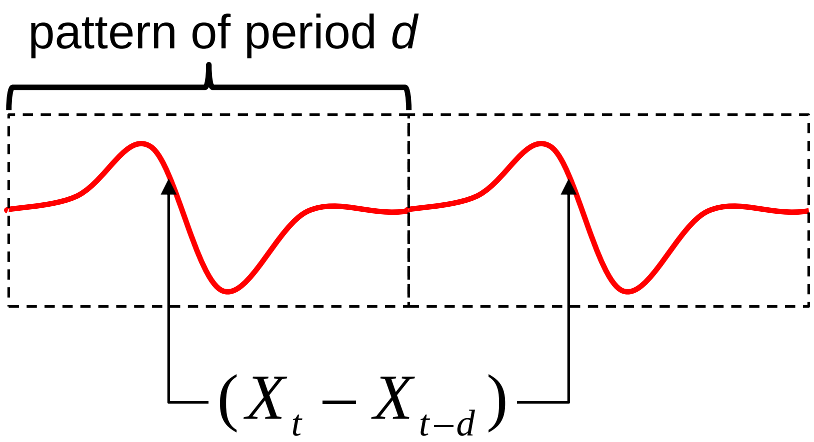


Differencing considerably decreases auto-correlations → random walk w/ drift is more plausible model for series

Seasonality

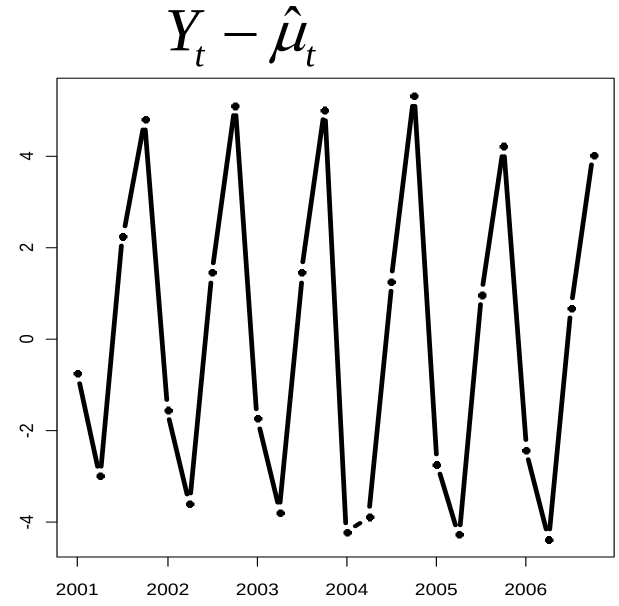
- Differencing can also be used for seasonal series $Y_t = S_t + X_t$
- If series has seasonal pattern with period d, then d-*lag* differences remove pattern
  - Define $Z_t = (1 - B^d) X_t$
    $= X_t - X_{t-d}$
  - If $S_t = S_{t+d}, \forall t \Rightarrow$

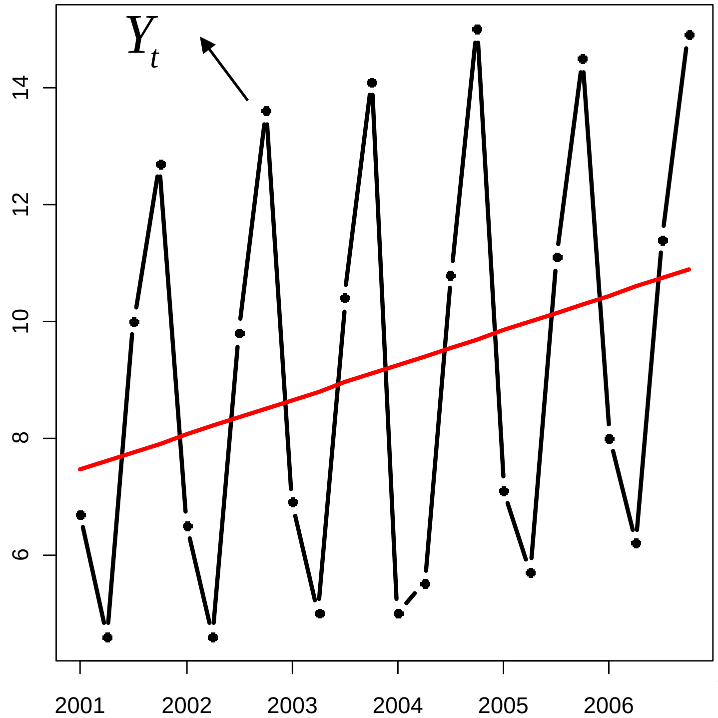
$\Rightarrow E[Z_t] = E[X_t - X_{t+d}] = 0$



Example

- Quarterly sales data
  (annual pattern→ d=4)





differencing also removed linear trend

Transformations

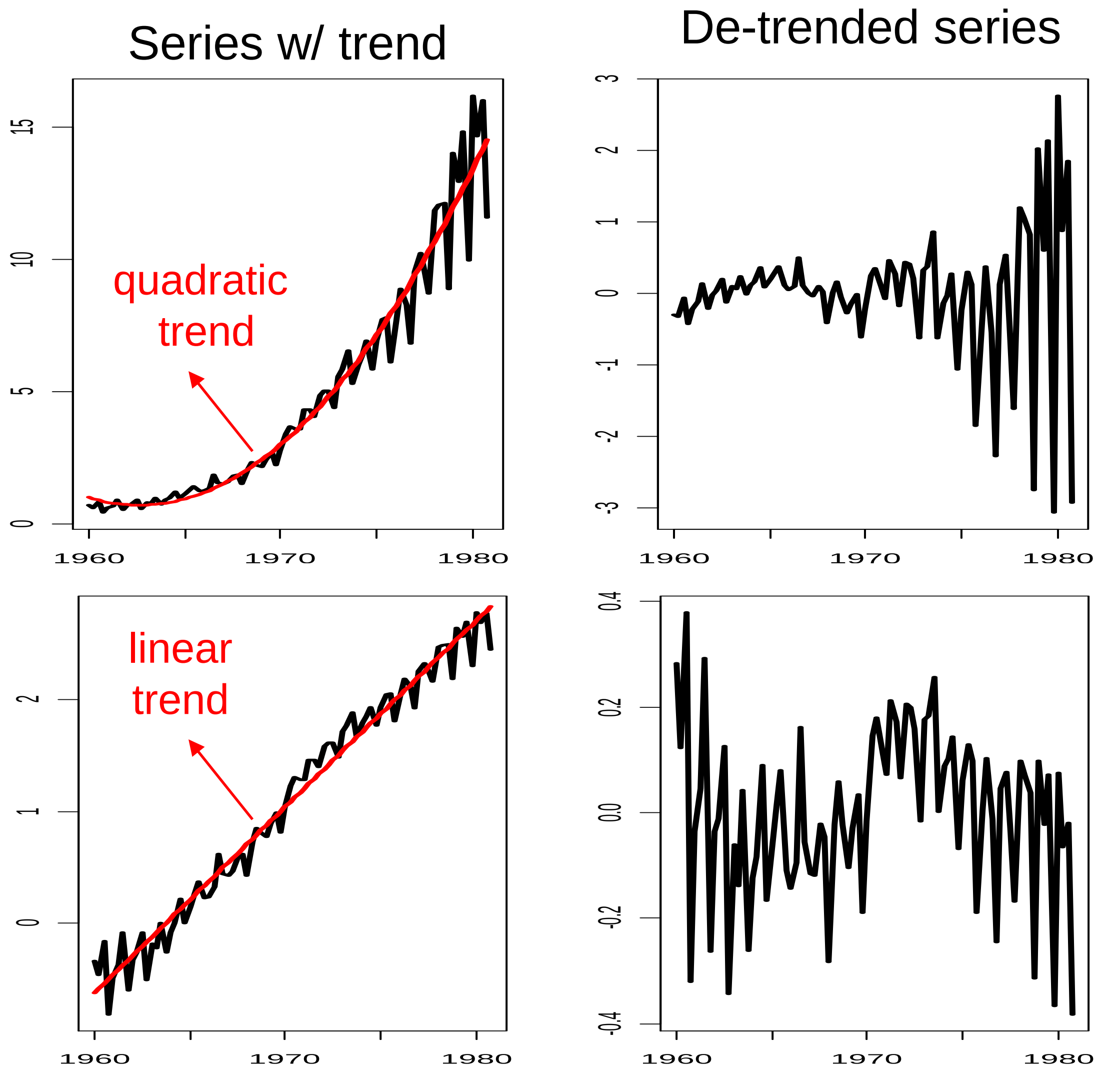
- If TS has non-constant variance, a nonlinear transformation can sometimes help
- For positive series in particular, the Box-Cox family of power transforms can be useful:

$$Y_t = \begin{cases} (X_t^{\lambda} - 1)/\lambda, & \lambda \neq 0 \\ \log(X_t), & \lambda = 0 \end{cases}$$

  - Try different values of λ, and check which one seems to give best results
  - Note: nonlinear transforms can also change μ_t

Example

- J&J sales
  - Original data:
  - Transformed data: $Y_t = \log(X_t)$

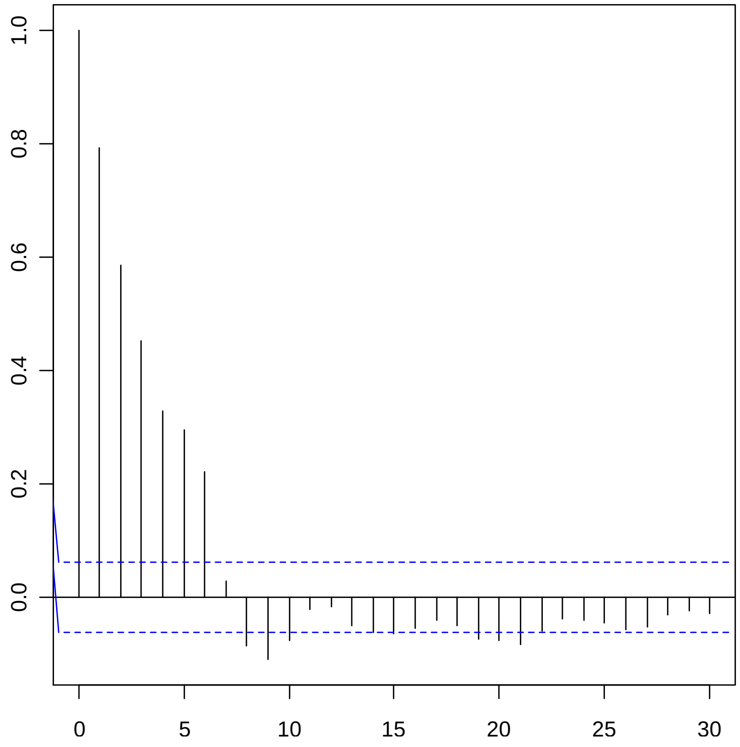


ACF Stationarity

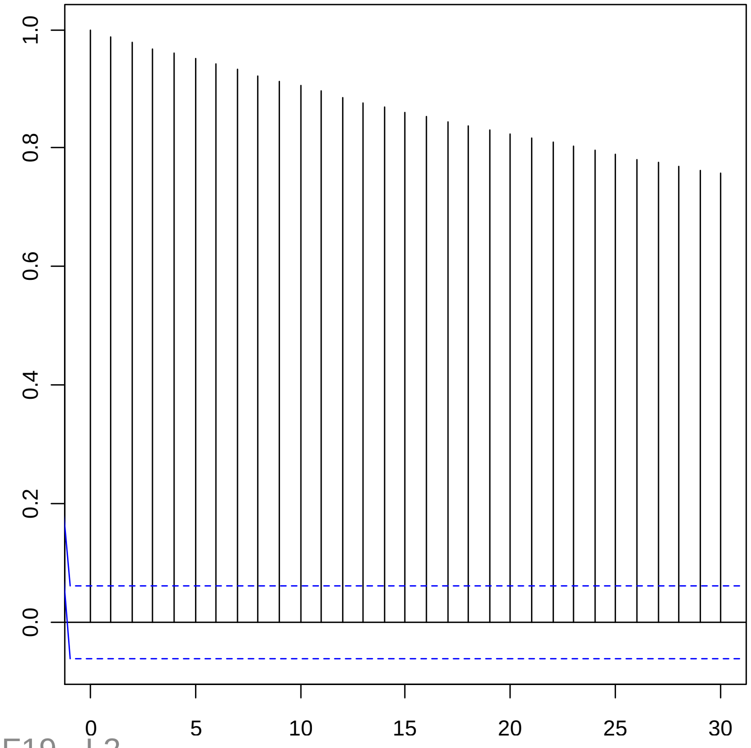
- Stationarity of $\gamma(s,t)$ is most difficult to check
- In practice, cannot verify stationarity since there is no way to estimate $\gamma(s,t)$, only $\gamma(h)$
  - Essentially, we "*assume*" stationarity in order to carry on with any TS analysis
- There are, however, 2 things we can check:
  - Sample ACF does not change when calculated from separate sub-sequences of data
  - Autocorrelations drop to zero relatively fast (i.e. exponentially) for large lags

Example

- Stationary ACF (AR model)

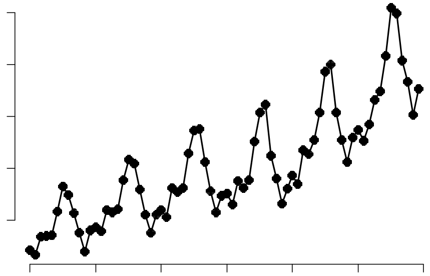


- Non-stationary ACF (Random walk)



TS Preprocessing Recap

- To make TS stationaty:
  - μ_t deterministic fn & σ_t^2 constant $\rightarrow$ detrending
    - estimate μ_t by MA smoother or regression
    - both trend & seasonality $\rightarrow$ decomposition
  - TS behaves like random walk (w/ polynomial drift) $\rightarrow$ differencing
  - μ_t & σ_t^2 deterministic fns $\rightarrow$ transformation
  - $\gamma(t,s)$ not stationary $\rightarrow$ differencing or other model



3. ARMA Models

STAD57 F19

Sotirios Damouras

Time Series Modeling

- Seen how to check for stationarity of TS data & how to “convert” them to stationarity if needed (de-trending, differencing, etc)
- Next step in analysis is to assume a *model* for the stationary part of the TS data
  - Need model to proceed with estimation / forecasting
  - Ideally, model should accurately describe TS dependence structure (i.e. sample ACF)
- For tractability, only consider *linear* models w.r.t. series’ lagged values (AR) and/or past WN (MA)

Stationary Linear Processes

- Important result (Wold's theorem) states that *every* zero-mean stationary TS can be expressed as:

$$X_t = \sum_{j=0}^{\infty} \psi_j W_{t-j}, \text{ where } \begin{cases} \sum_{j=0}^{\infty} \psi_j^2 < \infty \\ \{W_t\} \sim WN(0, \sigma_w^2) \end{cases}$$

  - $\{X_t\}$ is *1-sided linear* function of WN $\{W_t\}$
    - Process does *not depend on future* values of $\{W_t\}$
    - Only present & lagged values of $\{W_t\}$ appear in X_t
  - Square summability of ψ's ensures process is stable (variance does not explode to ∞)

Stationary Linear Processes

- Converse is also true, i.e. *every* Wold linear process is stationary
  - Proof: $X_t = \sum_{i=0}^{\infty} \psi_i \cdot W_{t-i}$, $W_t \sim WN(0, \sigma_w^2)$

$$\mu_t = \mathbb{E}[X_t] = \mathbb{E}\left[\sum_{i=0}^{\infty} \psi_i W_{t-i}\right] = \sum_{i=0}^{\infty} \psi_i \, \cancel{\mathbb{E}[W_{t-i}]}^{\,0} = 0$$

$$\sigma_t^2 = \mathbb{V}ar[X_t] = \mathbb{V}ar\left[\sum_{i=0}^{\infty} \psi_i \cdot W_{t-i}\right] = \sum_{i=0}^{\infty} \psi_i^2 \cdot \underbrace{\mathbb{V}ar[W_{t-i}]}_{=\sigma_w^2 \ \forall i} =$$

$$= \sum_{i=0}^{\infty} \psi_i^2 \cdot \sigma_w^2 = \sigma_w^2 \cdot \underbrace{\sum_{i=0}^{\infty} \psi_i^2}_{< \infty}$$

$$\gamma(s,t) = \gamma(s,s+h) = \mathrm{Cov}\left(\sum_{i=0}^{\infty} \psi_i \cdot W_{s-i}, \sum_{j=0}^{\infty} \psi_j \cdot W_{s+h-j}\right) =$$

$$= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_i \cdot \psi_j \cdot \underbrace{\mathrm{Cov}\left(W_{s-i}, W_{s+h-j}\right)}_{= \begin{cases} 0, & \text{if } s-i \neq s+h-j \\ \sigma_w^2, & \text{if } \cancel{s}-i = \cancel{s}+h-j \Rightarrow i = j-h \Rightarrow j = i+h \end{cases}} =$$

$$= \sum_{i=0}^{\infty} \psi_i \, \psi_{h+i} \cdot \sigma_w^2 = \sigma_w^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+h}, \quad \text{indep of } s,t$$

$$\Rightarrow \gamma(h) = \begin{cases} \sum_{i=0}^{\infty} \psi_i^2 \sigma_w^2, & \text{if } h = 0 \\ \left(\sum_{i=0}^{\infty} \psi_i \psi_{i+h}\right) \cdot \sigma_w^2, & \text{if } h \geq 1 \end{cases}$$

Time Series Modeling

So, why not model every stationary TS as stationary 1-sided linear process?

- Because model is *not* tractable in practice:
  - Need to estimate infinite # of parameters ψ_j
  - But only have finite # of data
    - Can only estimate ACF $\gamma(h)$ for $h=0,\ldots,n-1$
- Use simpler (finite) AR and/or MA models to describe TS dependence structure

Autoregressive Models

- Autoregressive (AR) models express current value as linear function of past values of TS
- AR model of order p, or AR(p), is of the form

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \cdots + \varphi_p X_{t-p} + W_t$$

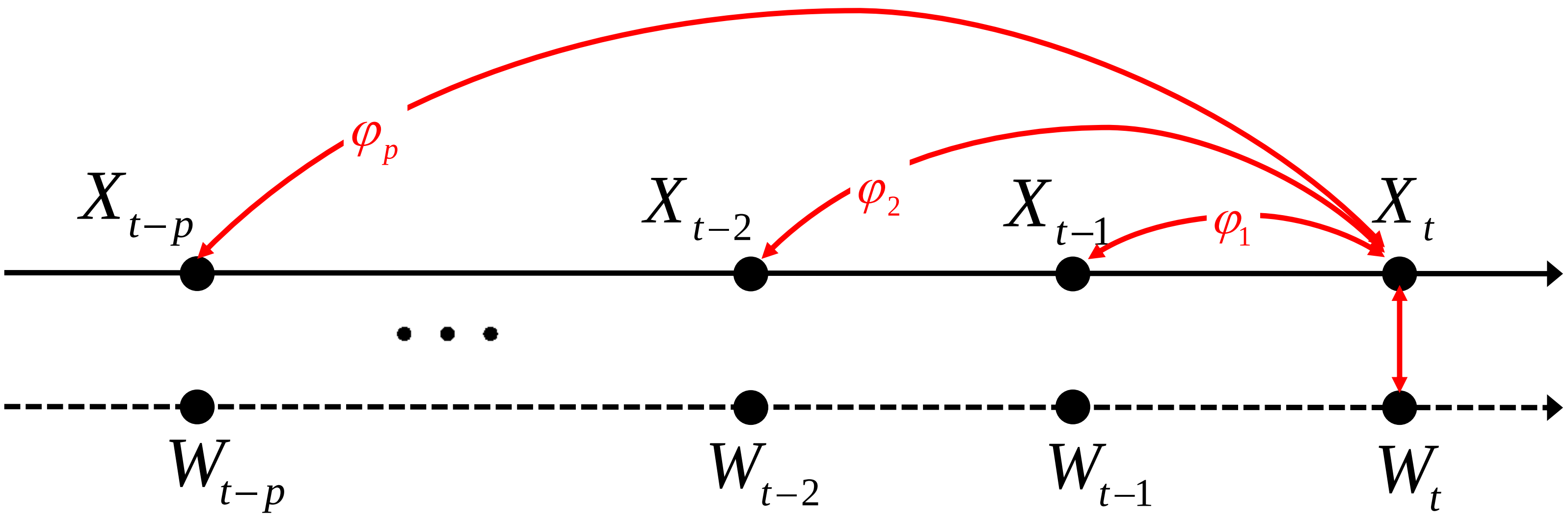
where $\varphi_1,\ldots, \varphi_p$ are constant, $\{W_t\}$ ~ WN($0,\sigma_w^2$), and $\{X_t\}$ is zero-mean

  - If $\mathrm{E}[X_t]=\mu\neq 0$, substitute $(X_t-\mu)$ in place of X_t:

$$(X_t - \mu) = \varphi_1 (X_{t-1} - \mu) + \cdots + \varphi_p (X_{t-p} - \mu) + W_t$$

Autoregressive Models

- $X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \cdots + \varphi_p X_{t-p} + W_t$



Autoregressive Operators

- Simplify representation of AR(p) model using *polynomials* of *backshift operator*
- The AR operator $\varphi(B)$ is defined as:

$$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p$$

  - Can write AR(p) model as: $\varphi(B) X_t = W_t \Leftrightarrow$

$$(1 - \varphi_1 B - \cdots - \varphi_p B^p) X_t = W_t \Leftrightarrow X_t - \varphi_1 X_{t-1} - \cdots - \varphi_p X_{t-p} = W_t$$

$$\Leftrightarrow X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + W_t$$

- Operator $\varphi(B)$ & corresponding polynomial are important for properties of AR model

Autoregressive Models

- *Not all* AR models are stationary (stable)
  - To check for stationarity, can express model as linear process $X_t = \sum_{j=0}^{\infty} \varphi_j W_{t-j}$ & check if its coefficients (φ_j) are square summable
- E.g. when is AR(1) model stationary?

$X_t = \varphi_1 X_{t-1} + W_t \Rightarrow X_t = \varphi_1 (\varphi_1 X_{t-2} + W_{t-1}) + W_t = \ldots = \sum_{j=0}^{\infty} (\varphi_1)^j W_{t-j}$

Want coefficient in Wold process representation to be square-summable $\Rightarrow \sum_{j=0}^{\infty} [(\varphi_1)^j]^2 = \sum_{j=0}^{\infty} (\varphi_1^2)^j < \infty \Rightarrow$

$\Rightarrow$ we need $\varphi_1^2 < 1 \Rightarrow |\varphi_1| < 1$ (in which case $\sum_{j\geq 0} (\varphi_1^2)^j = \frac{1}{1-\varphi_1^2}$)

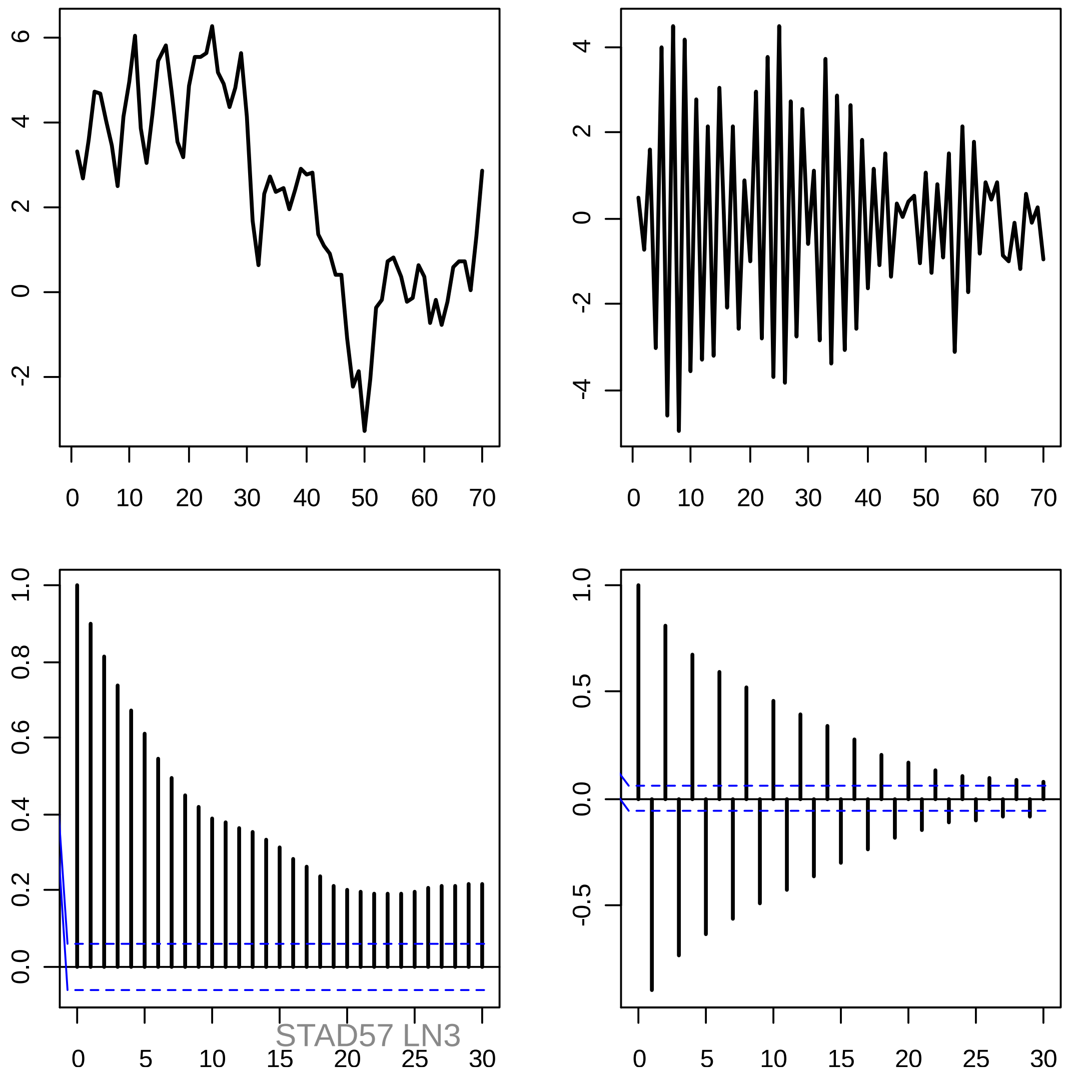
Example

$$X_t = .9X_{t-1} + W_t \qquad X_t = -.9X_{t-1} + W_t$$

Simulated series:

Sample ACF:

$\begin{pmatrix} \text{Theoretical} \\ \text{ACF: } \rho(h) = \varphi^h \end{pmatrix}$



Autoregressive Models

- For general AR(p) model, find linear representation through AR operator

  $\varphi(B)X_t = W_t$, where $\varphi(B) = 1 - \varphi_1 B - \cdots - \varphi_p B^p$

  - Try to find a polynomial operator $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$ so that: $X_t = \psi(B)W_t = \sum_{j=0}^{\infty} \psi_j W_{t-j} \quad (\psi_0 = 1)$

- To find $\psi(B)$, multiply both sides of $\varphi(B)X_t = W_t$ by $\psi(B)$: $\psi(B)\varphi(B)X_t = \psi(B)W_t = X_t \Rightarrow$

$$\Rightarrow \underline{\psi(B)\varphi(B) = 1}$$

Inverse Operators

- Solve for coefficients ψ_j of $\psi(B)$ so that:

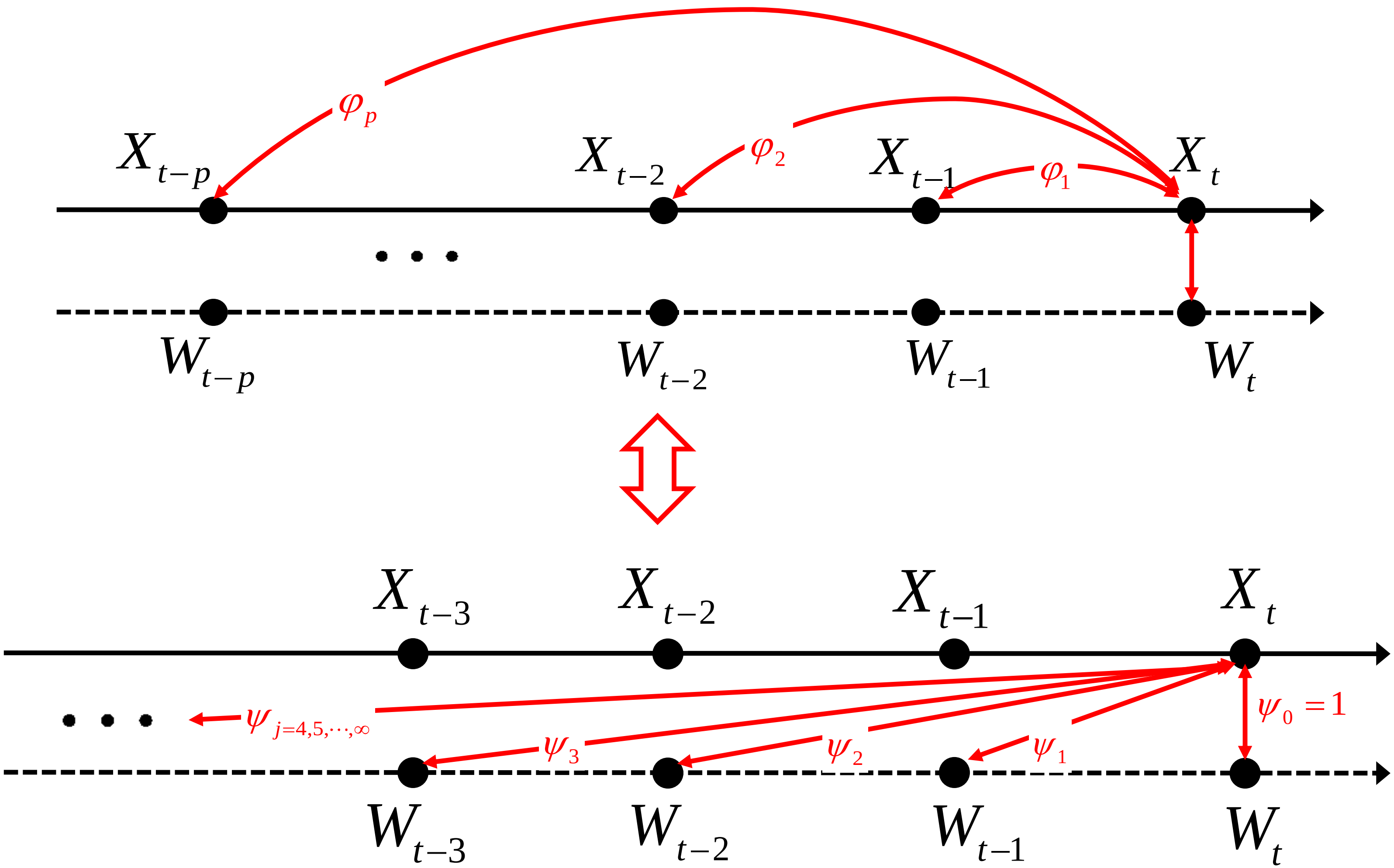
$$\psi(B)\varphi(B) = 1 \Leftrightarrow (1 + \psi_1 B + \psi_2 B^2 + \cdots)(1 - \varphi_1 B - \cdots - \varphi_p B^p) = 1$$

$$\Rightarrow 1 + (\psi_1 - \varphi_1)B + (\psi_2 - \varphi_1\psi_1 - \varphi_2)B^2 + \cdots = 1$$

$$\Rightarrow \begin{cases} \psi_1 - \varphi_1 = 0 \Rightarrow \psi_1 = \varphi_1 \\ \psi_2 - \varphi_1\psi_1 - \varphi_2 = 0 \Rightarrow \psi_2 = \varphi_1\psi_1 + \varphi_2 \\ \vdots \end{cases}$$

- Operator $\psi(B)$ is called the *inverse* of $\varphi(B)$, and is denoted by $\varphi^{-1}(B)$ or $1/\varphi(B)$

Autoregressive Models



Example

$$\sum_{j=0}^{\infty} \varphi^j \cdot W_{t-j}$$

- For AR(1) model $X_t = \varphi X_{t-1} + W_t \Leftrightarrow \varphi(B) X_t = W_t$, show that the inverse of $\varphi(B) = 1 - \varphi B$ is

$$\psi(B) = 1 + \varphi B + \varphi^2 B^2 + \varphi^3 B^3 + \cdots$$

want : $\psi(B) \cdot \varphi(B) = 1 \Rightarrow (1 + \psi_1 B + \psi_2 B^2 + \dots) \cdot (1 - \varphi B) = 1$

$\Rightarrow 1 + \underbrace{(\psi_1 - \varphi)}_{=0} \cdot B + \underbrace{(\psi_2 - \psi_1 \cdot \varphi)}_{=0} B^2 + \underbrace{(\psi_3 - \psi_2 \cdot \varphi)}_{=0} B^3 + \dots = 1$

$$\Rightarrow \begin{cases} \psi_1 - \varphi = 0 \\ \psi_2 - \psi_1 \varphi = 0 \\ \psi_3 - \psi_2 \varphi = 0 \\ \vdots \end{cases} \Rightarrow \begin{array}{l} \psi_1 = \varphi \\ \psi_2 = \varphi \cdot \psi_1 = \varphi^2 \\ \psi_3 = \varphi \cdot \psi_2 = \varphi^3 \\ \vdots \end{array}$$

$$\Rightarrow \begin{cases} \psi_j = \varphi^j, \ \forall j \geq 0 \\ \psi(B) = \varphi^{-1}(B) = \\ = 1 + \varphi B + \varphi^2 B^2 + \dots \end{cases}$$

Autoregressive Models

- Assume you express AR(p) model $\varphi(B)X_t = W_t$ as linear process $X_t = \varphi^{-1}(B)W_t = \psi(B)W_t$
  - Can you tell if $X_t = \psi(B)W_t$ is stationary?
- Important result: if AR(p) *characteristic polynomial,* $\varphi(z) = 1 - \varphi_1 z - \varphi_2 z^2 - \cdots - \varphi_p z^p$ for complex z, has *all* roots outside the unit circle, i.e. $\varphi(z) \neq 0$ for $|z| \leq 1$, then the inverse $\varphi^{-1}(B) = \psi(B)$ exists and has absolutely summable coefficients, i.e. $\sum_{j=0}^{\infty} |\psi_j| < \infty$

Causal Processes

- 1-sided linear process $X_t = \psi(B)W_t = \sum_{j=0}^{\infty} \psi_j W_{t-j}$ is called *causal* if $\sum_{j=0}^{\infty} |\psi_j| < \infty$
  - *Every causal* process is *stationary*, because

$$\sum_{j=0}^{\infty} |\psi_j| < \infty \Rightarrow \sum_{j=0}^{\infty} \psi_j^2 < \infty$$

- An AR(p) process is causal (→ stationary) if and only if the roots of its characteristic polynomial all lie outside the unit circle

$$\varphi(z) = 1 - \varphi_1 z - \varphi_2 z^2 - \cdots - \varphi_p z^p \neq 0 \text{ for all } |z| \leq 1$$

Example

$\varphi(B) \cdot X_t = W_t$

- Show that AR(1) model $X_t = \varphi X_{t-1} + W_t$ is causal (stationary) iff $|\varphi| < 1$

$\varphi(B) = 1 - \varphi B \Rightarrow$ characteristic polynomial is

$\varphi(z) = 1 - \varphi z$.

Root of $\varphi(z)$: $1 - \varphi z = 0 \Rightarrow \boxed{z = \frac{1}{\varphi}}$

The root is outside unit circle $\Leftrightarrow |z| > 1 \Leftrightarrow$

$\Leftrightarrow \left|\frac{1}{\varphi}\right| > 1 \Leftrightarrow |\varphi| < 1$

Moving Average Models

- Moving average (MA) models express TS as linear function of current & past values of WN
- MA model of order q, or MA(q), is of the form

$$X_t = W_t + \theta_1 W_{t-1} + \theta_2 W_{t-2} + \cdots + \theta_q W_{t-q}$$

  - where $\theta_1,\ldots,\theta_q$ are constant & $\{W_t\} \sim WN(0,\sigma_w^2)$

- Can write MA(q) using the MA operator θ(B)

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

Moving Average Models

- Every MA(q) process is stationary
  - Follows because it is *finite* linear process → coefficients are always summable
- However, can have two different MA(q) processes with the same autocovariance
  - MA models are not uniquely defined

Example

- Show that these MA models are equivalent

$$\begin{cases} \text{A: } X_t = W_t + \theta W_{t-1}, \text{ for } \{W_t\} \sim WN(0,1) \\ \text{B: } X_t = W_t + \frac{1}{\theta} W_{t-1}, \text{ for } \{W_t\} \sim WN(0,\theta^2) \end{cases}$$

both are MA(1)

For both models, $\mu_t = \mathbb{E}[X_t] = \mathbb{E}[W_t + \theta W_{t-1}] = 0 = \mathbb{E}[W_t + \frac{1}{\theta} W_{t-1}]$

$$\gamma_A(h) = \begin{cases} \text{for } h=0, & Cov(X_t, X_{t+h}) = Var[X_t] = Var[W_t + \theta W_{t-1}] = Var[W_t] + \theta^2 Var[W_{t-1}] = 1 + \theta^2 \\ \text{for } h=1, & Cov(X_t, X_{t+1}) = Cov(W_t + \theta W_{t-1}, W_{t+1} + \theta W_t) = \theta \cdot Cov(W_t, W_t) = \theta \\ \text{for } h \geq 2, & Cov(X_t, X_{t+h}) = \cdots = 0 \end{cases}$$

$$\gamma_B(h) = \begin{cases} \text{for } h=0, & Var(X_t] = Var[W_t + \frac{1}{\theta} W_{t-1}] = Var[W_t] + \frac{1}{\theta^2} Var[W_{t-1}] = \theta^2 + \frac{1}{\theta^2}\theta^2 = 1 + \theta^2 \\ \text{for } h=1, & Cov(X_t, X_{t+1}) = Cov(W_t + \frac{1}{\theta} W_{t-1}, W_{t+1} + \frac{1}{\theta} W_t) = \frac{1}{\theta} Var[W_t] = \frac{1}{\theta}\theta^2 = \theta \\ \text{for } h \geq 2, & Cov(X_t, X_{t+h}) = \cdots = 0 \end{cases}$$

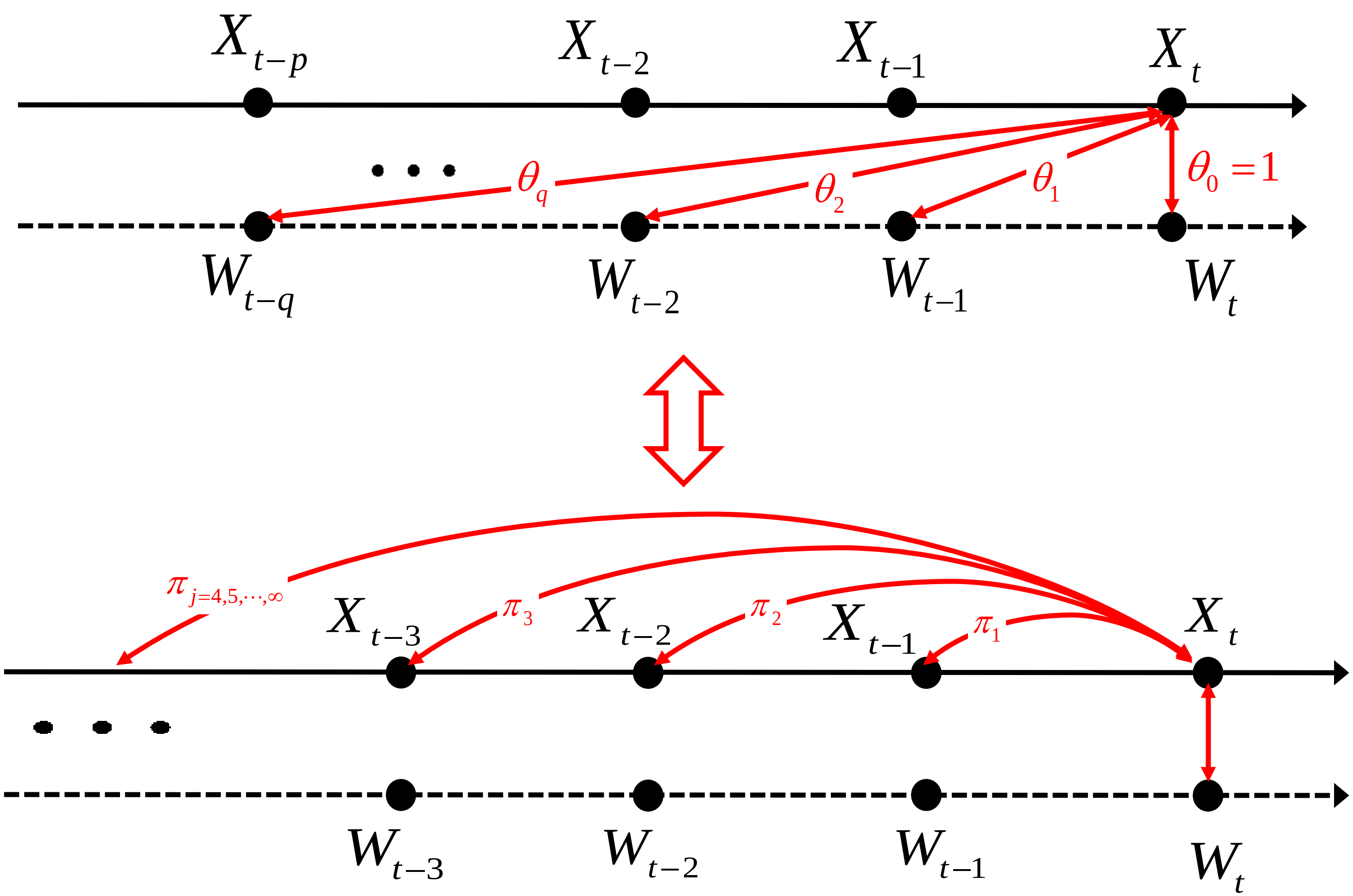
Invertible Models

- For MA(q) model $X_t = \theta(B)W_t$, use inverse operator of $\theta(B)$, call it $\pi(B)=\theta^{-1}(B)$, to express $\{X_t\}$ as *infinite* AR model

$$X_t = \theta(B)W_t \Rightarrow W_t = \pi(B)X_t = X_t + \pi_1 X_{t-1} + \pi_t X_{t-2} + \cdots$$

- To ensure AR(∞) representation is stable, we need π_j coefficients to be summable: $\sum_{j=0}^{\infty} \pi_j^2 < \infty$
  - Necessary & sufficient condition: MA characteristic equation has roots *outside* of unit disk $\theta(z) \neq 0, \ \forall |z| \leq 1$
  - Invertibility is useful: estimation algorithms rely on writing unobserved WN $\{W_t\}$ as function of TS $\{X_t\}$

Moving Average Models



Example

- MA models A & B are equivalent, but only A is invertible

$$\begin{cases} \text{A: } X_t = W_t + \underbrace{\tfrac{1}{2}}_{=\theta} W_{t-1}, \text{ for } \{W_t\} \sim WN(0,1) \\ \text{B: } X_t = W_t + 2 W_{t-1}, \text{ for } \{W_t\} \sim WN(0,(\tfrac{1}{2})^2) \end{cases}$$

$1/\theta$

(from $X_t = \theta(B) \cdot W_t$)

For A: $\theta(B) = 1 + \frac{1}{2} B \Rightarrow$ characteristic eqn: $\theta(z) = 1 + \frac{1}{2} \cdot z \Rightarrow$

$\Rightarrow$ root: $1 + \frac{1}{2} z = 0 \Rightarrow z = \frac{1}{1/2} = 2 \Rightarrow |z| = 2 > 1 \Rightarrow$ outside unit disc

$\Rightarrow$ invertible.

For B: $\theta(B) = 1 + 2 \cdot B \rightarrow$ characteristic eqn: $\theta(z) = 1 + 2 \cdot z \Rightarrow$

$\Rightarrow$ root: $1 + 2z = 0 \Rightarrow z = \frac{1}{2} \Rightarrow |z| = \frac{1}{2} \leq 1 \Rightarrow$ inside unit disc $\Rightarrow$

$\Rightarrow$ not invertible

ARMA models

- More general linear models arise from the combination of AR & MA models
- Autoregressive-Moving Average model of order (p,q), or just ARMA(p,q), is of the form:

$$X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + W_t + \theta_1 W_{t-1} + \cdots + \theta_q W_{t-q}$$

  - Can rewrite model using AR operator of order p & MA operator of order q, as: $\varphi(B)X_t = \theta(B)W_t$
  - If q=0 we just get AR(p) model, and if p=0 we get MA(q) model

Parameter Redundancy

- Certain ARMA(p,q) models are *redundant*, i.e. have more parameters than necessary
  - E.g. $X_t = .5X_{t-1} + W_t - .5W_{t-1}$ is simply $X_t = W_t$

$$\Rightarrow X_t - W_t = .5\cdot(X_{t-1} - W_{t-1})$$
$$= .5\,(.5\cdot(X_{t-2} - W_{t-2}))$$
$$= (.5)^2\cdot(.5\,(X_{t-3} - W_{t-3}))$$
$$\vdots$$
$$= (.5)^n\cdot(X_{t-n} - W_{t-n})$$

$\rightarrow 0$

and as $n \to \infty \Rightarrow X_t - W_t = \lim_{n\to\infty} (.5)^n\cdot(X_{t-n} - W_{t-n}) = 0 \Rightarrow$

$$\Rightarrow X_t - W_t = 0 \Rightarrow X_t = W_t$$

Parameter Redundancy

- Check for parameter redundancy using model's polynomial operators:
- ARMA model is redundant if $\varphi(B)$ & $\theta(B)$ have *common factors*.
  - E.g. $X_t = .5X_{t-1} + W_t - .5W_{t-1}$ $\Leftrightarrow \cancel{(1-.5B)}X_t = \cancel{(1-.5B)} \cdot W_t \Rightarrow$

    $\Rightarrow X_t = W_t$

- We will work with ARMA models in their simple (non-redundant) form, by simplifying any common factors in $\varphi(B)$ & $\theta(B)$

Example

$\varphi(B) = 1 - \varphi_1 B - \dots - \varphi_p B^p$

$\theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$ $\Leftrightarrow$

$\Leftrightarrow X_t = \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + W_t + \theta_1 W_{t-1} + \dots + \theta_q W_{t-q}$

- Find simple form of ARMA models
  - $X_t = .5X_{t-1} + W_t + .5W_{t-1}$ $\Leftrightarrow X_t - .5X_{t-1} = W_t + \frac{1}{2}W_{t-1} \Leftrightarrow$

φ, θ

$\Leftrightarrow (1 - \frac{1}{2}B)X_t = (1 + \frac{1}{2}B)W_t \Rightarrow$ no common factors $\Rightarrow$

$\Rightarrow$ model is already in simple form.

  - $X_t = -.5X_{t-1} + W_t - .25W_{t-2}$ $\longrightarrow$ $\begin{cases} \varphi(B) = 1 + \frac{1}{2}B \\ \theta(B) = 1 + 0B - \frac{1}{4}B^2 \end{cases}$

$\varphi(B)X_t = \theta(B)W_t \Rightarrow$

$\Rightarrow \cancel{(1 + \frac{1}{2}B)}X_t = (1 - \frac{1}{4}B^2)W_t =$

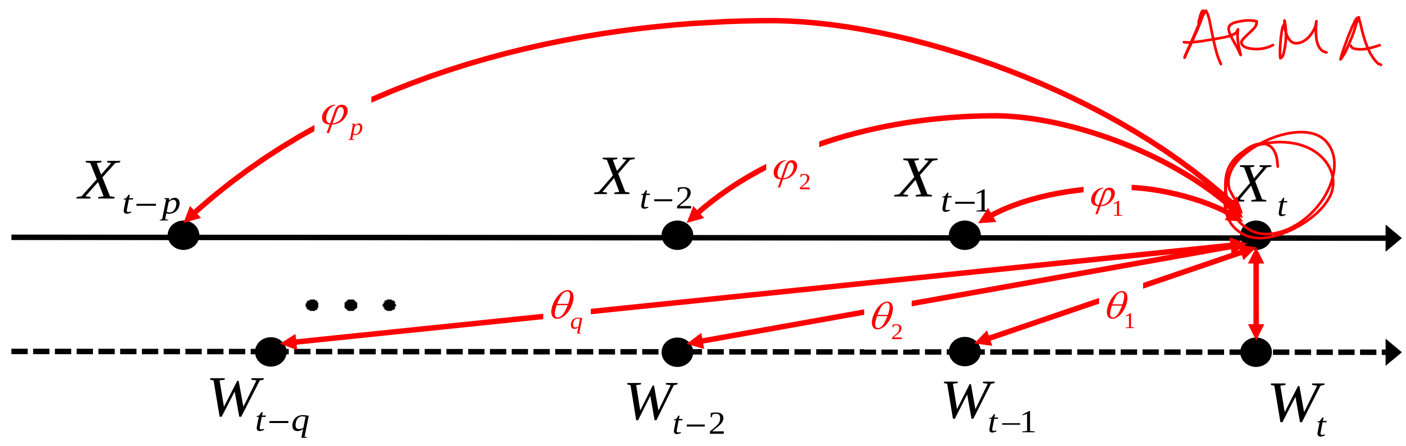
$= \cancel{(1 + \frac{1}{2}B)} \cdot (1 - \frac{1}{2}B)W_t \Leftrightarrow X_t = (1 - \frac{1}{2}B)W_t$

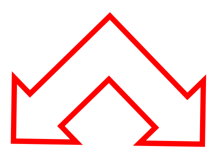
MA(1) $\longrightarrow$ $= W_t + \frac{1}{2}W_{t-1}$

ARMA Models

- For given ARMA model $\varphi(B)X_t = \theta(B)W_t \Leftrightarrow$
  $\Leftrightarrow X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + W_t + \theta_1 W_{t-1} + \cdots + \theta_q W_{t-q}$
  - Want to know if model is *causal*, i.e. can write $X_t = \psi(B)W_t = \sum_{j=0}^{\infty} \psi_j W_{t-j}$, for $\sum_{j=0}^{\infty} |\psi_j| < \infty$
  - Want to know if model is *invertible*, i.e. can write $W_t = \pi(B)X_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}$, for $\sum_{j=0}^{\infty} |\pi_j| < \infty$
- Important questions b/c causality implies stationarity & invertibility allows estimation
  - Can answer both using $\varphi(B)$ & $\theta(B)$

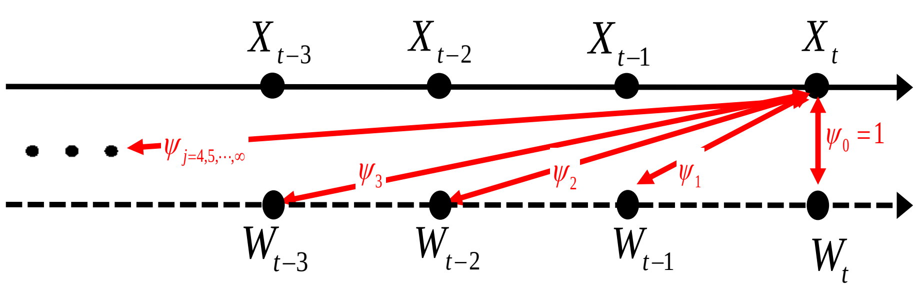
ARMA Models

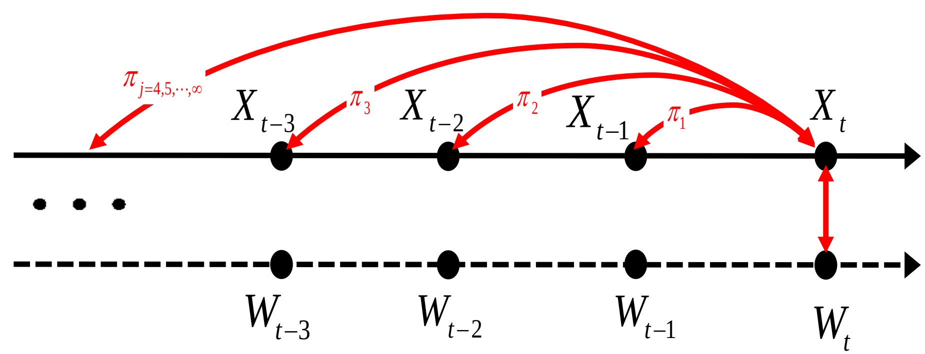




Causal Representation

Invertible Representation





Causality

- ARMA(p,q) process is causal (→ stationary) *if and only if* the roots of its AR characteristic polynomial all lie outside the unit circle:

$$\varphi(z) = 1 - \varphi_1 z - \varphi_2 z^2 - \cdots - \varphi_p z^p \neq 0 \text{ for all } |z| \leq 1$$

  - In this case, the inverse operator $\varphi^{-1}(B)$ exists and has absolutely summable coefficients
  - Causal representation of ARMA process is

$$\varphi(B) X_t = \theta(B) W_t \Rightarrow X_t = \varphi^{-1}(B)\theta(B) W_t = \psi(B) W_t$$

where $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j = \varphi^{-1}(B)\theta(B)$

Example

Want: $X_t = \psi(B) \cdot W_t$

$= [\varphi^{-1}(B)\,\theta(B)]\,W_t$

ARMA(1,1)

- Find causal representation of

$$X_t = .9X_{t-1} + W_t + .5W_{t-1} \Longrightarrow \begin{cases} \varphi(B) = 1 - .9B \\ \theta(B) = 1 + .5B \end{cases}$$

$$\psi(B) = \varphi^{-1}(B)\,\theta(B) = (1 + .9B + (.9)^2B^2 + \dots)\cdot(1 + .5B) =$$

$$= 1 + (.9+.5)B + .9\times(.9+.5)B^2 + (.9)^2\times(.9+.5)B^3 + \dots$$

$$= 1 + \overset{(.9)^0}{1.4}B + .9 \times 1.4\,B^2 + (.9)^2 \times 1.4\,B^3 + \dots$$

$$= 1 + 1.4 \cdot \sum_{j=1}^{\infty} (.9)^{j-1} \cdot B^j \Rightarrow$$

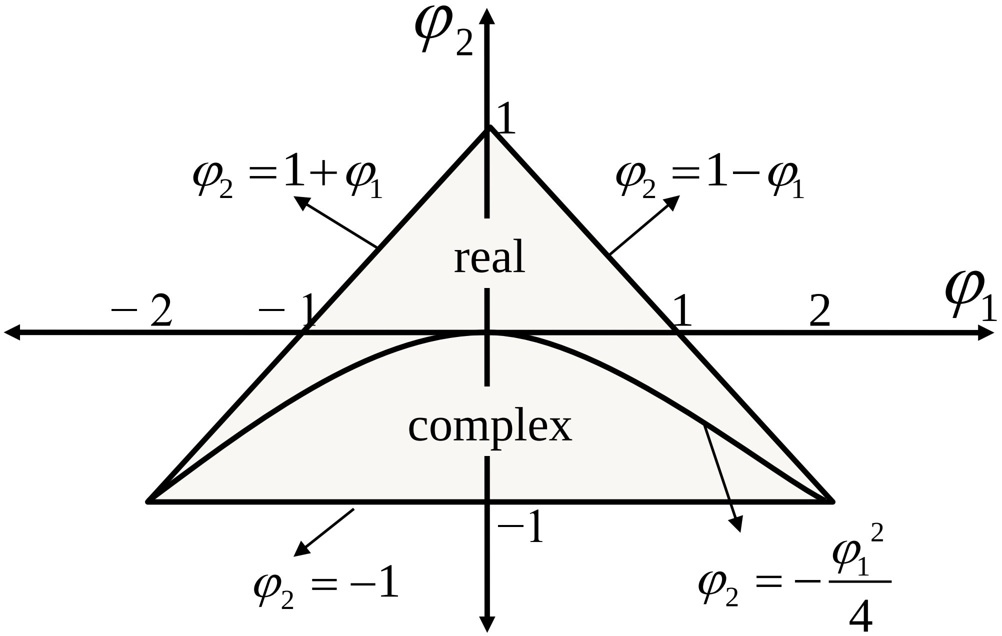
$$\Rightarrow X_t = W_t + 1.4 \cdot \sum_{j=1}^{\infty} (.9)^{j-1} \cdot W_{t-j}$$

Example

- ARMA(2,q) model $(1-\varphi_1 B-\varphi_2 B^2)X_t=\theta(B)W_t$ is causal, if parameters (φ_1, φ_2) satisfy:

$$\left|\frac{\varphi_1 \pm \sqrt{\varphi_1^2+4\varphi_2}}{-2\varphi_2}\right|>1 \Rightarrow$$

$$\Rightarrow \begin{cases} \varphi_1+\varphi_2<1 \\ \varphi_2-\varphi_1<1 \\ |\varphi_2|<1 \end{cases}$$



Invertibility

- ARMA(p,q) process is invertible *if and only if* the roots of its MA characteristic polynomial all lie outside the unit circle:

$$\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q \neq 0 \text{ for all } |z| \leq 1$$

  - In this case, the inverse operator $\theta^{-1}(\mathrm{B})$ exists and has absolutely summable coefficients
  - Invertible representation of ARMA process is

$$\varphi(B)X_t = \theta(B)W_t \Rightarrow W_t = \theta^{-1}(B)\varphi(B)X_t = \pi(B)X_t$$

where $\pi(B) = \sum_{j=0}^{\infty} \pi_j B^j = \theta^{-1}(B)\varphi(B)$

Example

ARMA(1,1)

- Find invertible representation of

$$X_t = .9X_{t-1} + W_t + .5W_{t-1} \Rightarrow \begin{cases} \varphi(B) = 1 - .9B \\ \theta(B) = 1 + .5B = 1 - (-.5)B \end{cases}$$

$$\Rightarrow \theta^{-1}(B) = \sum_{j=0}^{\infty} (-.5)^j \cdot B^j = 1 - \frac{1}{2}B + \frac{1}{4}B^2 - \frac{1}{8}B^3 + \ldots$$

$$\Rightarrow \pi(B) = \theta^{-1}(B)\,\varphi(B) = \left(1 - \frac{1}{2}B + \frac{1}{4}B^2 - \frac{1}{8}B^3 + \ldots\right) \cdot (1 - .9B) =$$

$$= 1 - (.9 + .5) \cdot B + (.5) \cdot (.9 + .5)B^2 - (.5)^2(.9 + .5)B^3 + \ldots$$

$$= 1 - 1.4 \cdot \sum_{j=1}^{\infty} (-.5)^{j-1} \cdot B^j \Rightarrow$$

$$\Rightarrow W_t = X_t - 1.4X_{t-1} + 1.4(.5)X_{t-2} - 1.4(.5)^2 X_{t-3} + \ldots$$

ARMA Model

- Will be using ARMA models to describe times series dynamics: $\varphi(B)X_t = \theta(B)W_t \Leftrightarrow$

$$\Leftrightarrow X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + W_t + \theta_1 W_{t-1} + \cdots + \theta_q W_{t-q}$$

  - Model must be *causal* (i.e. stationary) & *invertible* $\leftrightarrow$ all roots of $\varphi(z)=0$ & $\theta(z)=0$ have $|z|>1$
- Next, want to know the dependence structure (i.e. ACF) of any ARMA model
  - Want to match that to sample ACF from data

ACF of ARMA Model

- Since ARMA model is stationary, could use its causal representation to find ACF
  - For causal process $X_t = \psi(B)W_t = \sum_{j=0}^{\infty} \psi_j W_{t-j}$, ACF is given by: $\rho(h) = \left(\sum_{j=0}^{\infty} \psi_j \psi_{j+h}\right) \Big/ \left(\sum_{j=0}^{\infty} \psi_j^2\right)$
  - Find ψ-weights from: $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j = \varphi^{-1}(B)\theta(B)$
    - If model has MA part *only* → *finite* # of ψ-weights which are given by θ(B): $\psi(B) = \theta(B) \Rightarrow \psi_j = \theta_j$
    - If model has AR part → *infinite* # of ψ-weights
      - In practice, since weights are absolutely summable (go to 0 exponentially fast), can *approximate* ρ(h) using only a large but finite number of ψ's

Finding ψ-weights

$\psi(B) = \varphi^{-1}(B) \cdot \theta(B)$

- For general ARMA(p,q): $\varphi(B)\psi(B) = \theta(B) \Rightarrow$

$$(1-\varphi_1 B - \cdots - \varphi_p B^p)(\psi_0 + \psi_1 B + \psi_2 B^2 + \cdots) = (1 + \theta_1 B + \cdots + \theta_q B^q)$$

Solve by matching coefficients of powers of B:

$$\begin{aligned}
\psi_0 &= 1 \\
\psi_1 - \varphi_1\psi_0 &= \theta_1 \\
\psi_2 - \varphi_1\psi_1 - \varphi_2\psi_0 &= \theta_2 \\
&\vdots \\
\psi_{p-1} - \varphi_1\psi_{p-2} - \cdots - \varphi_{p-1}\psi_0 &= \theta_{p-1} \\
\psi_p - \varphi_1\psi_{p-1} - \cdots - \varphi_p\psi_0 &= \theta_p \\
&\vdots \\
\psi_j - \varphi_1\psi_{j-1} - \cdots - \varphi_p\psi_{j-p} &= 0 \\
&\vdots
\end{aligned}$$

First p equations have different # of terms

$\theta_j = 0$, for $j > q$

Remaining equations have same # of terms

Finding ψ-weights

- Rewrite last equations more succinctly as:

$$\begin{cases} \psi_j - \sum_{k=1}^{j} \varphi_k \psi_{j-k} = \theta_j, \text{ for } 0 \le j < \max(p, q+1) \\ \psi_j - \sum_{k=1}^{p} \varphi_k \psi_{j-k} = 0, \text{ for } j \ge \max(p, q+1) \end{cases}$$

where $\varphi_j = 0$ for $j > p$ & $\theta_j = 0$ for $j > q$

- Equations can be solved by substitution:

$$\underline{\psi_0 = 1}$$

$$\psi_1 - \varphi_1 \psi_0 = \theta_1 \Rightarrow \underline{\psi_1 = \varphi_1 + \theta_1}$$

$$\psi_2 - \varphi_1 \psi_1 - \varphi_2 \psi_0 = \theta_2 \Rightarrow \underline{\psi_2 = \varphi_1(\varphi_1 + \theta_1) + \varphi_2 + \theta_2}$$

$$\vdots$$

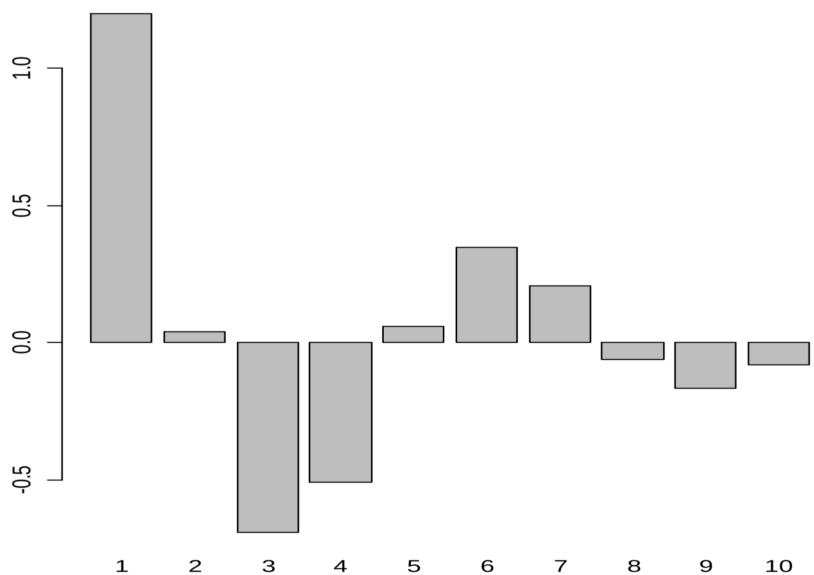
Finding ψ-weights in R

- R function `ARMAtoMA()` calculates ψ-weights
  - E.g. for $X_t = .7X_{t-1} - .3X_{t-2} + W_t + .5W_{t-1} - .2W_{t-2}$

list of φ's
list of θ's
# of ψ's

```
ARMAtoMA(ar=c(.7,-.3), ma=c(.5,-.2), lag.max=10)
```

Resulting ψ-weights:
(first 10 lags)



ACF of ARMA Model

- There is easier way to find ACF of ARMA model, *without* having to calculate all the ψ's & then using them to find causal model's ACF
- Trick is to find & solve a recurrence equation (like the one for ψ's) *directly* in terms of $\rho(h)$
- Begin with the simpler case of a pure AR process, and then move to general ARMA

ACF of AR Model

$X_t \to \mu_t = 0$
$\to \sigma_t^2 = \text{constant}$

- Consider AR(p) model: $X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + W_t$
  - Take equation for X_{t+h} & multiply both sides by X_t

$\to X_{t+h} = \varphi_1 X_{t+h-1} + \cdots + \varphi_p X_{t+h-p} + W_{t+h} \Rightarrow$

$\Rightarrow X_{t+h} \cdot X_t = \varphi_1 X_{t+h-1} \cdot X_t + \cdots + \varphi_p X_{t+h-p} \cdot X_t + W_{t+h} \cdot X_t$

  - Now take expectations & divide both sides by $\gamma(0) = \sigma_t^2$ to get: $\rho(h) = \varphi_1 \rho(h-1) + \cdots + \varphi_p \rho(h-p)$

$\Rightarrow \mathbb{E}[X_{t+h} \cdot X_t] = \varphi_1 \mathbb{E}[X_{t+h-1} \cdot X_t] + \cdots + \varphi_p \mathbb{E}[X_{t+h-p} \cdot X_t] + \mathbb{E}[W_{t+h} \cdot X_t]$

$\Rightarrow Cov(X_{t+h}, X_t) = \varphi_1 Cov(X_{t+h-1}, X_t) + \cdots + \varphi_p Cov(X_{t+h-p}, X_t) + Cov(W_{t+h}, X_t)$

$\Rightarrow \frac{\gamma(h)}{\gamma(0)} = \varphi_1 \frac{\gamma(h-1)}{\gamma(0)} + \cdots + \varphi_p \frac{\gamma(h-p)}{\gamma(0)} + 0$

(b/c $Cov(W_{t+h}, X_t) = Cov(W_{t+h}, \sum_{j=0}^{\infty} \psi_j W_{t-j}) = \sum_{j=0}^{\infty} \psi_j Cov(W_{t+h}, W_{t-j}) \to 0$)

$\Rightarrow \rho(h) = \varphi_1 \rho(h-1) + \cdots + \varphi_p \rho(h-p)$

ACF of AR Model

- Want to solve *recurrence* (or *difference*) *equation*: $\rho(h) = \varphi_1 \rho(h-1) + \cdots + \varphi_p \rho(h-p),$ for $h \geq 1$
  - Get first p values $\{\rho(0),...,\rho(p-1)\}$ (*initial conditions*) & use substitution into equation for $h=p,p+1,\ldots$
- To get initial conditions, use ρ(0)=1 & the recurrence equations for $h=1,...,p-1$,, together with the fact that $\rho(-h)=\rho(h)$ for any $h\geq 1$
  - This gives set of #p linear eqn's with #p unknowns, which can be solved to get $\{\rho(0),...,\rho(p-1)\}$

Example

AR(2), w/ $\phi(B) = 1 - .5B + .3B^2$

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

w/ $\phi_1 = .5$, $\phi_2 = -.3$

- Find $\rho(h)$ for h=1,2,3 of AR: $X_t = .5X_{t-1} - .3X_{t-2} + W_t$

ϕ_1 ϕ_2

$$\rho(h) = \phi_1 \rho(h-1) + \phi_2 \rho(h-2), \quad \forall h$$

with initial conditions
$$\begin{cases} \rho(0) = 1 \\ \rho(1) = \phi_1 \cdot \rho(1-1) + \phi_2 \cdot \rho(1-2) = \phi_1 \rho(0) + \phi_2 \rho(-1) \end{cases}$$

$\rho(0) = 1$, $\rho(-1) = \rho(1)$ $\Rightarrow$

$$\Rightarrow \begin{cases} \rho(0) = 1 \\ \rho(1) = \phi_1 + \phi_2 \cdot \rho(1) \end{cases} \Rightarrow \begin{cases} \rho(0) = 1 \\ \rho(1) = \dfrac{\phi_1}{1-\phi_2} = \dfrac{.5}{1-(-.3)} = \dfrac{.5}{1.3} \end{cases}$$

$$\rho(2) = \phi_1 \rho(1) + \phi_2 \rho(0) = .5 \times \frac{.5}{1.3} - .3 \cdot 1$$

$$\rho(3) = \phi_1 \rho(2) + \phi_2 \rho(1) = .5 \times \left[.5 \times \frac{.5}{1.3} - .3\right] - .3 \cdot \left(\frac{.5}{1.3}\right]$$

ACF of ARMA Model

- For general ARMA $X_t = \sum_{j=1}^{p} \varphi_j X_{t-j} + \sum_{j=0}^{q} \theta_j W_{t-j}$ $(\theta_0 = 1)$ recurrence equation for γ(h) is given by:

  $$\gamma(h) = \varphi_1 \gamma(h-1) + \cdots + \varphi_p \gamma(h-p), \text{ for } h \geq \max(p, q+1)$$

  with initial conditions:

$$\gamma(h) = \sum_{j=1}^{p} \varphi_j \gamma(h-j) + \sigma_w^2 \sum_{j=h}^{q} \theta_j \psi_{j-h}, \text{ for } 0 \leq h < \max(p, q+1)$$

  - Initial conditions define system of linear equations, which can be solved to find $\{\gamma(0),...,\gamma(\max(p-1,q))\}$
  - Once we know $\gamma(0)$, we can find $\rho(\mathrm{h})=\gamma(\mathrm{h})/\gamma(0)$
  - We only need to know ψ_j for $0 \leq j \leq q$ $(\psi_0=1)$

ACF of ARMA Model

- Derivation of recurrence equation

$$X_{t+h} = \sum_{j=1}^{p} \varphi_j X_{t+h-j} + \sum_{j=0}^{q} \theta_j W_{t+h-j} \Rightarrow \text{(multiply by } X_t\text{)}$$

$$\Rightarrow X_{t+h} \cdot X_t = \sum_{j=1}^{p} \varphi_j X_{t+h-j} X_t + \sum_{j=0}^{q} \theta_j W_{t+h-j} \cdot X_t \Rightarrow \text{(take } \mathbb{E}\text{)}$$

$$\Rightarrow \underbrace{\mathbb{E}[X_{t+h} X_t]}_{=\gamma(h)} = \sum_{j=1}^{p} \varphi_j \underbrace{\mathbb{E}[X_{t+h-j} \cdot X_t]}_{=\gamma(h-j)} + \sum_{j=0}^{q} \theta_j \mathbb{E}[W_{t+h-j} \cdot X_t]$$

Write $\mathbb{E}[X_t \cdot W_{t+h-j}] = \mathbb{E}\left[\left(\sum_{k=0}^{\infty} \psi_k W_{t-k}\right) \cdot W_{t+h-j}\right] = \sum_{k=0}^{\infty} \psi_k \cdot \mathbb{E}[W_{t-k} \cdot W_{t+h-j}]$.

where $\mathbb{E}[W_{t-k} \cdot W_{t+h-j}] = \sigma_w^2$ only if $t-k = t+h-j \Rightarrow k = j-h$ $\left(\mathbb{E}[W_{t-k} W_{t+h-j}] = 0 \text{ if } k \neq j-h\right)$

$\Rightarrow \mathbb{E}[X_t \cdot W_{t+h-j}] = \psi_{j-h} \cdot \sigma_w^2$. Also note that if $h \geq q+1 \Rightarrow$

$\Rightarrow j-h < 0 \Rightarrow \psi_{j-h} = 0, \ \forall j-h < 0$ b/c $\mathbb{E}[W_{t-k} \cdot W_{t+h-j}] = 0, \ \forall k \geq 0, \ 0 \leq j \leq q$

Putting results together, we have:

$$\gamma(h) = \sum_{j=1}^{p} \varphi_j \gamma(h-j) + \sigma_w^2 \cdot \sum_{j=0}^{q} \theta_j \psi_{j-h} \quad \left(\text{where } \psi_{j-h}=0 \text{ when } j-h<0\right), \quad \forall\, h \geq 0$$

$$\Rightarrow \begin{cases} \text{initial conds:} & \gamma(h) = \sum_{j=1}^{p} \varphi_j \gamma(h-j) + \sigma_w^2 \cdot \sum_{j=h}^{q} \theta_j \psi_{j-h}, \\ & \forall\, 0 \leq h < \max(p, q+1) \\ \text{recurrence eqn:} & \gamma(h) = \sum_{j=1}^{p} \varphi_j \gamma(h-j), \quad \forall\, h \geq \max(p, q+1) \end{cases}$$

Example

$\varphi(B) = 1 - \varphi B \Rightarrow \varphi^{-1}(B) = \sum_{j=0}^{\infty} \varphi^j B^j$

$\psi(B) = \varphi^{-1}(B) \cdot \theta(B) = (1 + \varphi B + \varphi B^2 + \dots) \cdot (1 + \theta B)$

$= \psi_0 + \psi_1 \cdot \underline{\underline{B}} + \psi_2 \cdot B^2 + \dots \Rightarrow \psi_1 = \varphi + \theta$

- Find $\rho(h)$ for general ARMA(1,1) model:

$$X_t = \varphi X_{t-1} + W_t + \theta W_{t-1}$$

$p = q = 1 \Rightarrow \max(p, q+1) = 2 \Rightarrow$

$\Rightarrow \begin{cases} \text{initial conds:} & \gamma(h) = \varphi \cdot \gamma(h-1) + \sigma_w^2 \cdot \sum_{j=h}^{q} \theta_j \psi_{j-h}, \text{ for } h = 0, 1 \\ \text{recurrence equ:} & \gamma(h) = \varphi \gamma(h-1), \quad \forall h \geq 2 \end{cases}$

initial conds: $\begin{cases} \gamma(0) = \varphi \cdot \gamma(0-1) + \sigma_w^2 \cdot \sum_{j=0}^{1} \theta_j \cdot \psi_{j-0} & (h=0) \\ \gamma(1) = \varphi \gamma(1-1) + \sigma_w^2 \sum_{j=1}^{1} \theta_j \psi_{j-1} & (h=1) \end{cases} \Leftrightarrow$

$\Leftrightarrow \begin{cases} \gamma(0) = \varphi \cdot \gamma(1) + \sigma_w^2 (\theta_0 \cdot \psi_0 + \theta_1 \cdot \psi_1) \\ \gamma(1) = \varphi \gamma(0) + \sigma_w^2 \theta_1 \cdot \psi_0 \end{cases}$, where $\begin{cases} \theta_0 = \psi_0 = \varphi_0 = 1 \\ \theta_1 = \theta, \quad \varphi_1 = \varphi \\ \psi_1 = \varphi + \theta \end{cases}$

$$\Leftrightarrow \begin{cases} \gamma(0) = \varphi\gamma(1) + \sigma_w^2 \cdot [1 + \theta\cdot(\varphi+\theta)] \\ \gamma(1) = \varphi\gamma(0) + \sigma_w^2 \cdot \theta \end{cases} \Rightarrow$$

$$\Rightarrow \gamma(0) = \varphi[\varphi\gamma(0) + \sigma_w^2\theta] + \sigma_w^2 \cdot [1 + \theta\varphi + \theta^2] \Rightarrow$$

$$\Rightarrow \boxed{\gamma(0) = \sigma_w^2 \cdot \frac{(1+2\varphi\theta+\theta^2)}{1-\varphi^2}} \rightarrow \text{Variance of ARMA(1,1)}$$

$$\gamma(1) = \varphi\gamma(0) + \sigma_w^2\theta = \varphi\cdot\sigma_w^2 \cdot \frac{1+2\varphi\theta+\theta^2}{1-\varphi^2} + \sigma_w^2\theta \Rightarrow$$

$$\Rightarrow \boxed{\gamma(1) = \sigma_w^2 \cdot \frac{(\varphi+\theta)\cdot(1+\varphi\theta)}{1-\varphi^2}}$$

And from recurrence

$$\gamma(2) = \varphi\gamma(1)$$

$$\gamma(3) = \varphi\gamma(2) = \varphi^2\gamma(1)$$

$$\vdots$$

$$\gamma(h) = \varphi^{h-1}\cdot\gamma(1)$$

$$\Rightarrow \gamma(h) = \varphi^{h-1}\cdot\sigma_w^2 \cdot \frac{(\varphi+\theta)(1+\varphi\theta)}{1-\varphi^2}, \forall h \geq 1$$

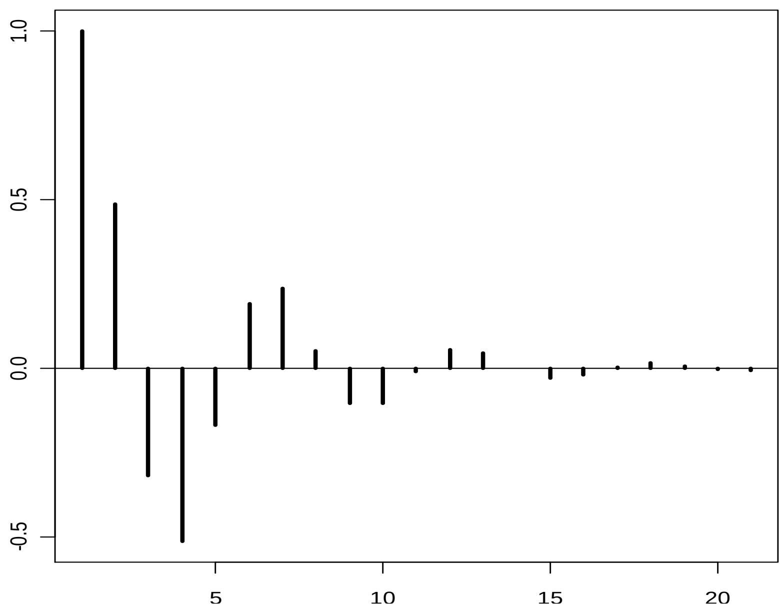
$$\Rightarrow \rho(h) = \frac{\gamma(h)}{\gamma(0)} = \varphi^{h-1} \frac{(\varphi+\theta)\cdot(1+\varphi\theta)}{1+2\varphi\theta+\theta^2}, \forall h \geq 1$$

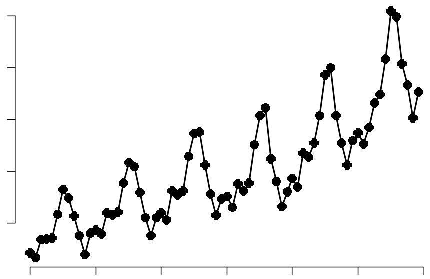
Finding ACF in R

- R function `ARMAacf()` calculates $\rho(h)$
  - E.g. for $X_t = .7X_{t-1} - .3X_{t-2} + W_t + .5W_{t-1} - .2W_{t-2}$

$$\texttt{ARMAacf(ar=c}\overbrace{\texttt{(.7,-.3)}}^{\text{list of } \varphi\text{'s}}\texttt{, ma=c}\overbrace{\texttt{(.5,-.2)}}^{\text{list of } \theta\text{'s}}\texttt{, lag.max=}\overbrace{\texttt{20}}^{\#\text{ of } \rho\text{'s}}\texttt{)}$$

Resulting ACF:
(first 20 lags)





4. ARMA Prediction

STAD57 F19

Sotirios Damouras

Forecasting

- Have seen how to use:
  - ACF to describe TS dependence structure
  - ARMA for modeling stationary TS
- Next, look at *forecasting* (or prediction)
  - Goal is to predict future values of a TS based on its current & past values
  - Assume *stationary* TS model with *known* parameters
    - Will return to parameter estimation later (first need some forecasting results)

Minimum MSE Predictors

- For stationary TS $\{X_t\}$, assume you observe $\mathbf{X}=\{X_1,\ldots,X_n\}$ and want to predict X_{n+m}
  - Denote the *m-step-ahead predictor* by X_{n+m}^{n}
- *Minimum mean square error* (MMSE) predictor is *conditional expectation* of $X_{n+m} \mid \mathbf{X}$

$$\text{MMSE}: \; X_{n+m}^{n} = \mathbb{E}\left[X_{n+m} \mid \mathbf{X}\right] = \mathbb{E}\left[X_{n+m} \mid X_1,\ldots,X_n\right]$$

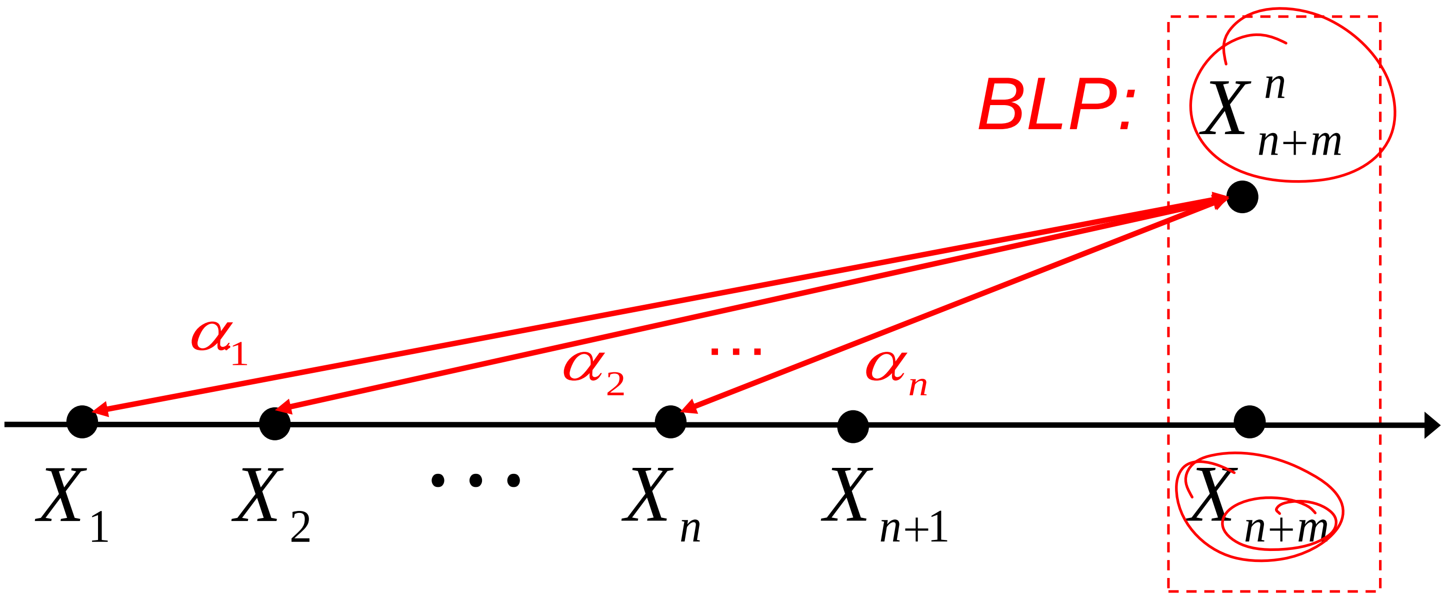
  - MMSE is function g() of **X** that minimizes MSE:

$$\mathbb{E}\left[\left(X_{n+m} - g(\mathbf{X})\right)^2\right]$$

Best Linear Predictors

- MMSE is difficult to find in general
  - Depends on entire joint distr. of $\{\mathbf{X}, X_{n+m}\}$
- For simplicity, only focus on *linear* predictors of the form: $X_{n+m}^{n} = g(\mathbf{X}) = \alpha_0 + \sum_{k=1}^{n} \alpha_k X_k$
- *Best linear predictor* (BLP) is *linear* predictor that minimizes the MSE $\mathbb{E}[(X_{n+m} - g(\mathbf{X}))^2]$
  - BLP's depend only on 2nd order moments of $\{X_t\}$ (i.e. covariances) which are easy to estimate
  - If $\{X_t\}$ is Gaussian, then BLP=MMSE predictor

Best Linear Predictors



(MMSE $\mathbb{E}[X_{n+m} \mid \mathbf{X}]$ can be nonlinear fn of $\{X_1, \ldots, X_n\}$)

Best Linear Predictors

- The coefficients $\alpha_0,\ldots,\alpha_n$ of the BLP

  $X_{n+m}^{n} = \alpha_0 + \sum_{k=1}^{n} \alpha_k X_k = \sum_{k=0}^{n} \alpha_k X_k$ (where $X_0 = 1$)

  are found by solving the set of equations:

  $$\mathbb{E}\left[\left(X_{n+m} - X_{n+m}^{n}\right) X_k\right] = 0, \forall k = 0,\ldots,n$$

  - Proof: $MSE = \mathbb{E}\left[(X_{n+m} - X_{n+m}^{n})^2\right]$. To minimize the MSE, take derivatives w.r.t. the coefficients α_k & set to 0.

$$\frac{d}{d\alpha_k} \mathbb{E}\left[\left(X_{n+m} - \sum_{j=0}^{n} \alpha_j X_j\right)^2\right] = \mathbb{E}\left[\frac{d}{d\alpha_k}\left(X_{n+m} - \sum_{j=0}^{n} \alpha_j X_j\right)^2\right] =$$

$$= \mathbb{E}\left[2 \cdot \Big(X_{n+m} - \underbrace{\sum_{j=0}^{n} \alpha_j X_j}_{= X^n_{n+m}} \Big) \cdot \frac{d}{d\alpha_k} \Big(X_{n+m} - \sum_{j=0}^{n} \alpha_j X_j \Big) \right] =$$

$$= \cancel{2} \cdot \mathbb{E}\left[\left(X_{n+m} - X^n_{n+m} \right) \cdot X_k \right] = 0 \ , \ \forall \ k = 0, \dots, n$$

Best Linear Predictors

- Solving for α_0, we have: $\alpha_0 = \mu - \mu \underbrace{\sum_{k=1}^{n} \alpha_k}_{=1}$
  - Proof: for $k=0$, we have $\mathbb{E}\left[\left(X_{n+m} - X_{n+m}^{n}\right) \cdot \cancel{X_0}\right] = 0 \Rightarrow$

$$\Rightarrow \mathbb{E}\left[X_{n+m} - \sum_{j=0}^{n} \alpha_j X_j\right] = 0 \Rightarrow \underbrace{\mathbb{E}\left[X_{n+m}\right]}_{=\mu} = \sum_{k=0}^{n} \alpha_k \underbrace{\mathbb{E}\left[X_k\right]}_{=\mu,\ \forall k \neq 0} \Rightarrow$$

$$\Rightarrow \mu = \alpha_0 \cdot 1 + \sum_{j=1}^{n} \alpha_j \cdot \mu \Rightarrow \alpha_0 = \mu - \mu \cdot \sum_{j=1}^{n} \alpha_j$$

- So, we have: $X_{n+m}^{n} = \left(\mu - \mu \sum_{k=1}^{n} \alpha_k\right) + \sum_{k=1}^{n} \alpha_k X_k \Leftrightarrow$

$$\Leftrightarrow \left(X_{n+m}^{n} - \mu\right) = \sum_{k=1}^{n} \alpha_k \left(X_k - \mu\right)$$

  - For simplicity & without loss of generality, assume that $\{X_t\}$ is zero-mean, so that $\alpha_0 = 0$

Best Linear Predictors

- Let $\{X_t\}$ be 0-mean & stationary, and consider the 1-step-ahead BLP: $X_{n+1}^n = \sum_{k=1}^n \alpha_k X_k \Leftrightarrow$ $X_{n+1}^n = \varphi_{n1} X_n + \varphi_{n2} X_{n-1} + \cdots + \varphi_{nn} X_1$, where $\underline{\alpha_k = \varphi_{n,n+1-k}}$
- The coefficients $\{\varphi_{n1}, \ldots, \varphi_{nn}\}$ satisfy:

$$\sum_{j=1}^n \varphi_{nj} \gamma(k-j) = \gamma(k), \ \forall k = 1, \ldots, n$$

- Proof: We know that BLP must satisfy the "normal equations" $\mathbb{E}\left[(X_{n+1} - X_{n+1}^n) \cdot X_k\right] = 0, \ \forall \ k = 1, \ldots, n$

$$\Rightarrow \mathbb{E}\Big[\Big(X_{n+1} - \overbrace{\sum_{j=1}^{n} \varphi_{n,n-j+1} \cdot X_j}^{X_{n+1}^n}\Big) \cdot X_k\Big] = 0 \quad \Rightarrow$$

$$\Rightarrow \underbrace{\mathbb{E}[X_{n+1} \cdot X_k]} - \sum_{j=1}^{n} \varphi_{n,n-j+1} \underbrace{\mathbb{E}[X_j X_k]} = 0$$

$$\Leftrightarrow Cov(X_{n+1}, X_k) = \sum_{j=1}^{n} \varphi_{n,n-j+1} \, Cov(X_j, X_k)$$

$$\Leftrightarrow \gamma(n+1-k) = \sum_{j=1}^{n} \varphi_{n,n-j+1} \cdot \gamma(j-k), \quad \forall\, k = 1, \dots, n$$

$$\Leftrightarrow \gamma(h) = \sum_{j=1}^{n} \varphi_{n,j}\, \gamma(h-j), \quad \text{for } h = n+1-k, \; k = 1, \dots, n$$

$$\Rightarrow h = 1, \dots, n$$

Best Linear Predictors

$$\sum_{j=1}^{n} \varphi_{nj}\gamma(k-j) = \gamma(k), \ \ \forall k = 1,\ldots,n \Leftrightarrow$$

$$\Leftrightarrow \left\{ \begin{array}{c} (k=1) \ \ \varphi_{n1}\gamma(1-1) + \varphi_{n2}\gamma(1-2) + \cdots + \varphi_{nn}\gamma(1-n) = \gamma(1) \\ (k=2) \ \ \varphi_{n1}\gamma(2-1) + \varphi_{n2}\gamma(2-2) + \cdots + \varphi_{nn}\gamma(2-n) = \gamma(2) \\ \vdots \\ (k=n) \ \ \varphi_{n1}\gamma(n-1) + \varphi_{n2}\gamma(n-2) + \cdots + \varphi_{nn}\gamma(n-n) = \gamma(n) \end{array} \right\}$$

$$\Leftrightarrow \left\{ \begin{array}{c} \varphi_{n1}\gamma(0) + \varphi_{n2}\gamma(1) + \cdots + \varphi_{nn}\gamma(n-1) = \gamma(1) \\ \varphi_{n1}\gamma(1) + \varphi_{n2}\gamma(0) + \cdots + \varphi_{nn}\gamma(n-2) = \gamma(2) \\ \vdots \\ \varphi_{n1}\gamma(n-1) + \varphi_{n2}\gamma(n-2) + \cdots + \varphi_{nn}\gamma(0) = \gamma(n) \end{array} \right\}$$

Best Linear Predictors

- System can be written in matrix form as:

$$\underbrace{\begin{bmatrix} \gamma(0) & \gamma(1) & \gamma(2) & \cdots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \gamma(1) & \cdots & \gamma(n-2) \\ \gamma(2) & \gamma(1) & \gamma(0) & \cdots & \gamma(n-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma(n-1) & \gamma(n-2) & \gamma(n-3) & \cdots & \gamma(0) \end{bmatrix}}_{=\mathbf{\Gamma}_n} \times \underbrace{\begin{bmatrix} \varphi_{n1} \\ \varphi_{n2} \\ \varphi_{n3} \\ \vdots \\ \varphi_{nn} \end{bmatrix}}_{=\boldsymbol{\varphi}_n} = \underbrace{\begin{bmatrix} \gamma(1) \\ \gamma(2) \\ \gamma(3) \\ \vdots \\ \gamma(n) \end{bmatrix}}_{=\boldsymbol{\gamma}_n} \Leftrightarrow$$

$$\Leftrightarrow \underline{\mathbf{\Gamma}_n \boldsymbol{\varphi}_n = \boldsymbol{\gamma}_n}$$

Best Linear Predictors

- BLP coefficients φ_{nj} are given by:

$$\mathbf{\Gamma}_n \boldsymbol{\varphi}_n = \boldsymbol{\gamma}_n \Rightarrow \boldsymbol{\varphi}_n = \mathbf{\Gamma}_n^{-1} \boldsymbol{\gamma}_n$$

  - Matrix $\mathbf{\Gamma}_n$ is *symmetric* & *positive definite* (provided γ(0)>0) → it is *invertible*, and we can solve for $\boldsymbol{\varphi}_n$
  - 1-step-ahead BLP can be written as:

$$X_{n+1}^{n} = \sum_{j=1}^{n} \varphi_{nj} X_{n+1-j} = \boldsymbol{\varphi}_n' \mathbf{X}$$

where $\mathbf{X} = [X_n \quad X_{n-1} \quad \cdots \quad X_1]'$

Best Linear Predictors

- MSE P_{n+1}^{n} of 1-step-ahead BLP is given by:

$$P_{n+1}^{n} = \mathbb{E}[(X_{n+1} - X_{n+1}^{n})^2] = \gamma(0) - \boldsymbol{\gamma}_n' \boldsymbol{\Gamma}_n^{-1} \boldsymbol{\gamma}_n$$

$\varphi_n' = \gamma_n' \cdot \Gamma_n^{-1}$

$\underline{X} = \begin{bmatrix} X_n \\ \vdots \\ X_1 \end{bmatrix}$

- Proof:

$$P_{n+1}^{n} = \mathbb{E}[(X_{n+1} - \underbrace{X_{n+1}^{n}}_{\varphi_n' \cdot \underline{X}})^2] = \mathbb{E}[(X_{n+1} - \varphi_n' \cdot \underline{X})^2] =$$

$$= \mathbb{E}[(X_{n+1} - \gamma_n' \cdot \Gamma_n^{-1} \cdot \underline{X})^2] = \mathbb{E}[X_{n+1}^2 - 2 \cdot \gamma_n' \Gamma_n^{-1} \underline{X} \cdot X_{n+1} +$$

$$+ \gamma_n' \Gamma_n^{-1} \underline{X} \cdot \underline{X}' \cdot \Gamma_n^{-1} \cdot \gamma_n] =$$

$$= \underbrace{\mathbb{E}[X_{n+1}^2]}_{} - 2 \gamma_n' \Gamma_n^{-1} \cdot \mathbb{E}[\underline{X} \cdot X_{n+1}] + \gamma_n' \Gamma_n^{-1} \mathbb{E}[\underline{X} \cdot \underline{X}'] \cdot \Gamma_n^{-1} \gamma_n$$

$$= \gamma(0)$$

$$\mathbb{E}[\underline{X} \cdot X_{n+1}] = \mathbb{E}\left\{ \begin{bmatrix} X_n \\ X_{n-1} \\ \vdots \\ X_1 \end{bmatrix} \cdot X_{n+1} \right\} = \begin{bmatrix} \mathbb{E}[X_n X_{n+1}] \\ \vdots \\ \mathbb{E}[X_1 X_{n+1}] \end{bmatrix} = \begin{bmatrix} \gamma(1) \\ \vdots \\ \gamma(n) \end{bmatrix} = \underline{\gamma}_n$$

$$\mathbb{E}[\underline{X} \cdot \underline{X}'] = \mathbb{E}\left\{ \begin{bmatrix} X_n \\ \vdots \\ X_1 \end{bmatrix} \cdot [X_n \cdots X_1] \right\} = \mathbb{E}\left\{ \begin{bmatrix} X_n^2 & X_n X_{n-1} & \cdots & X_n \cdot X_1 \\ X_{n-1}X_n & X_{n-1}^2 & \cdots & X_{n-1} X_1 \\ \vdots & \vdots & \ddots & \vdots \\ X_1 X_n & X_1 X_{n-1} & \cdots & X_1^2 \end{bmatrix} \right\} =$$

$$= \begin{bmatrix} \mathbb{E}[X_n^2] & \mathbb{E}[X_n X_{n-1}] & \mathbb{E}[X_n \cdot X_1] \\ \mathbb{E}[X_{n-1} X_n] & \mathbb{E}[X_{n-1}^2] & \cdots \mathbb{E}[X_{n-1} X_1] \\ \vdots & & \vdots \\ \mathbb{E}[X_1 X_n] & \mathbb{E}[X_1 X_{n-1}] & \cdots \mathbb{E}[X_1^2] \end{bmatrix} = \begin{bmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(n-1) & \gamma(n-2) & \cdots & \gamma(0) \end{bmatrix} = \Gamma_n$$

$$\Rightarrow P_{n+1}^n = \gamma(0) - 2 \cdot \underline{\gamma}_n' \cdot \Gamma_n^{-1} \cdot \underbrace{\mathbb{E}[\underline{X} X_{n+1}]}_{= \underline{\gamma}_n} + \underline{\gamma}_n' \Gamma_n^{-1} \underbrace{\mathbb{E}[\underline{X}\,\underline{X}']}_{= \Gamma_n} \Gamma_n^{-1} \underline{\gamma}_n =$$

$$= \gamma(0) - 2 \underline{\gamma}_n' \Gamma_n^{-1} \underline{\gamma}_n + \underline{\gamma}_n' \cancel{\Gamma_n^{-1}} \cancel{\Gamma_n} \Gamma_n^{-1} \underline{\gamma}_n =$$

$$= \gamma(0) - \underline{\gamma}_n' \Gamma_n^{-1} \underline{\gamma}_n$$

Example

$\Gamma_n \varphi_n = \gamma_n \iff \varphi_n = \Gamma_n^{-1} \cdot \gamma_n$

- Consider AR(2) model $X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + W_t$
  - Find $\boldsymbol{\varphi}_n$ for n=1,2,3,…

For n=1 : Want to find $\varphi_1 = \varphi_{11}$ s.t. $\underbrace{\gamma(0)}_{\Gamma_1} \cdot \varphi_{11} = \underbrace{\gamma(1)}_{\gamma_1} \Rightarrow \varphi_{11} = \frac{\gamma(1)}{\gamma(0)} = \rho(1) = \frac{\varphi_1}{1-\varphi_2}$

For AR(2): $\rho(h) = \varphi_1 \rho(h-1) + \varphi_2 \rho(h-2)$, $\forall h \geq 2$,

with initial conditions: $\begin{cases} \rho(0) = 1 \\ \rho(1) = \varphi_1 \rho(0) + \varphi_2 \rho(1) \end{cases}$

$\Rightarrow \begin{cases} \rho(0) = 1 \\ \rho(1) = \varphi_1 \cdot + \varphi_2 \cdot \rho(1) \Rightarrow \rho(1) = \frac{\varphi_1}{1-\varphi_2} \end{cases}$

For n=2 : Want to find $\varphi_2 = \begin{bmatrix} \varphi_{21} \\ \varphi_{22} \end{bmatrix}$ s.t. $\Gamma_2 \varphi_2 = \gamma_2 \Rightarrow$

$\Rightarrow \begin{bmatrix} \gamma(0) & \gamma(1) \\ \gamma(1) & \gamma(0) \end{bmatrix} \cdot \begin{bmatrix} \varphi_{21} \\ \varphi_{22} \end{bmatrix} = \begin{bmatrix} \gamma(1) \\ \gamma(2) \end{bmatrix} \Rightarrow \begin{bmatrix} \varphi_{21} \\ \varphi_{22} \end{bmatrix} = \frac{1}{\gamma^2(0) - \gamma^2(1)} \cdot \begin{bmatrix} \gamma(0) & -\gamma(1) \\ -\gamma(1) & \gamma(0) \end{bmatrix} \cdot \begin{bmatrix} \gamma(1) \\ \gamma(2) \end{bmatrix} \Rightarrow$

$$\Rightarrow \begin{cases} \varphi_{21} = \dfrac{\gamma(0)\gamma(1) - \gamma(1)\gamma(2)}{\gamma^2(0) - \gamma^2(1)} \\[2ex] \varphi_{22} = \dfrac{-\gamma^2(1) + \gamma(0)\gamma(2)}{\gamma^2(0) - \gamma^2(1)} \end{cases}$$

divide both numerator & denominator by $\gamma^2(0)$ $\Rightarrow$

$$\begin{cases} \varphi_{21} = \dfrac{\rho(1) - \rho(1)\rho(2)}{1 - \rho^2(1)} \\[2ex] \varphi_{22} = \dfrac{\rho(2) - \rho^2(1)}{1 - \rho^2(1)} \end{cases}$$

We have: $\begin{cases} \rho(1) = \dfrac{\varphi_1}{1-\varphi_2} \\ \rho(2) = \varphi_1\rho(1) + \varphi_2\rho(0) \end{cases}$ for φ_2

$\xrightarrow{\text{subst. in}}$ eqn

$$\begin{cases} \varphi_{21} = \ldots = \varphi_1 \\ \varphi_{22} = \ldots = \varphi_2 \end{cases}$$

For $n=3$, can show $\varphi_3 = \begin{bmatrix} \varphi_{31} \\ \varphi_{32} \\ \varphi_{33} \end{bmatrix} = \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ 0 \end{bmatrix}$

For $n=4$, can show $\varphi_4 = \begin{bmatrix} \varphi_{41} \\ \varphi_{42} \\ \varphi_{43} \\ \varphi_{44} \end{bmatrix} = \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ 0 \\ 0 \end{bmatrix}$

etc.

Best Linear Predictors

- For AR(p) model $X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + W_t$

  and for $n \geq p$, we have: $\begin{cases} \varphi_{nj} = \varphi_j, & j = 1, \ldots, p \\ \varphi_{nj} = 0, & j = p+1, \ldots, n \end{cases}$

  - To prove this, just write $X_{n+1}^n = \sum_{j=1}^{p} \varphi_j X_{n+1-j}$
  - This is the BLP because:

$= \sum_{j=1}^{p} \varphi X_{n+1-j} + W_{n+1}$

$$\mathbb{E}\left[\left(X_{n+1} - X_{n+1}^n\right) X_k\right] = \mathbb{E}\left[\left(X_{n+1} - \sum_{j=1}^{p} \varphi_j X_{n+1-j}\right) X_k\right] =$$

$$= \mathbb{E}\left[W_{n+1} X_k\right] = 0, \ \forall k = 1, \ldots, n$$

Durbin-Levinson Algorithm

- To find $\boldsymbol{\varphi}_n$ for general ARMA model, have to solve linear system: $\boldsymbol{\Gamma}_n \boldsymbol{\varphi}_n = \boldsymbol{\gamma}_n \Rightarrow \boldsymbol{\varphi}_n = \boldsymbol{\Gamma}_n^{-1} \boldsymbol{\gamma}_n$
  - For large n, this can be very time consuming (need to invert $n \times n$ matrix)
- Fortunately, there is an *iterative* algorithm for solving the system which is a lot faster
  - Algorithm takes advantage of *special structure* of $\boldsymbol{\Gamma}_n$, which is *symmetric* with *equal diagonal elements* (a.k.a. Toeplitz matrix)

$$\boldsymbol{\Gamma}_n = \begin{bmatrix} \gamma(0) & \gamma(1) & \gamma(2) & \cdots \\ \gamma(1) & \gamma(0) & \gamma(1) & \ddots \\ \gamma(2) & \gamma(1) & \gamma(0) & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{bmatrix}$$

Durbin-Levinson Algorithm

- The BLP equations: $\begin{cases} \boldsymbol{\varphi}_n = \boldsymbol{\Gamma}_n^{-1}\boldsymbol{\gamma}_n \\ P_{n+1}^n = \gamma(0) - \boldsymbol{\gamma}_n'\boldsymbol{\Gamma}_n^{-1}\boldsymbol{\gamma}_n \end{cases}$

  can be solved iteratively (in n), as follows:

Start with $\varphi_{00} = 0$ & $P_1^0 = \gamma(0)$, and for $n \ge 1$ set:

$$\varphi_{nn} = \frac{\gamma(n) - \sum_{k=1}^{n-1} \varphi_{n-1,k}\gamma(n-k)}{\gamma(0) - \sum_{k=1}^{n-1} \varphi_{n-1,k}\gamma(k)} = \frac{\rho(n) - \sum_{k=1}^{n-1} \varphi_{n-1,k}\rho(n-k)}{1 - \sum_{k=1}^{n-1} \varphi_{n-1,k}\rho(k)}$$

$$P_{n+1}^n = P_n^{n-1}\left(1 - \varphi_{nn}^2\right)$$

$$\varphi_{nk} = \varphi_{n-1,k} - \varphi_{nn}\varphi_{n-1,n-k}, \quad k = 1, 2, \ldots, n-1 \ \text{(if } n \ge 2\text{)}$$

- Algorithm progresses as: $\boldsymbol{\varphi}_0 \rightarrow \boldsymbol{\varphi}_1 \rightarrow \boldsymbol{\varphi}_2 \rightarrow \boldsymbol{\varphi}_3 \rightarrow \ldots$

Example

DL algo $\begin{cases} \varphi_{nn} = \dfrac{\rho(n) - \sum_{k=1}^{n-1} \varphi_{n-1,k} \cdot \rho(n-k)}{\rho(0) - \sum_{k=1}^{n-1} \varphi_{n-1,k} \cdot \rho(k)} \\ P_{n+1}^{n} = P_n^{n-1} \cdot (1 - \varphi_{nn}^2) \end{cases}$

$\varphi_{n,k} = \varphi_{n-1,k} - \varphi_{nn} \cdot \varphi_{n-1,n-k}, \ \forall\, k = n-1, \dots, 1$

- Use the Durbin-Levinson algorithm to find $\boldsymbol{\varphi}_3$ & P_4^3 based on $\gamma(0),\ \rho(1), \rho(2), \rho(3)$

$(n=0)$ $\varphi_{00} = 0$, $P_1^0 = \gamma(0)$

$(n=1)$ $\varphi_{11} = \dfrac{\rho(1)}{\rho(0)} = \rho(1)$, $P_2^1 = P_1^0 \cdot (1 - \rho^2(1))$

$(n=2)$ $\varphi_{22} = \dfrac{\rho(2) - \varphi_{11} \cdot \rho(1)}{\rho(0) - \varphi_{11} \cdot \rho(1)}$

$P_3^2 = P_2^1 (1 - \varphi_{22}^2) = \gamma(0) \cdot (1 - \varphi_{11}^2)(1 - \varphi_{22}^2)$

$\varphi_{21} = \varphi_{11} - \varphi_{22} \cdot \varphi_{11} = \varphi_{11} \cdot (1 - \varphi_{22})$

For $n=3$

$$\varphi_{33} = \frac{\rho(3) - \sum_{k=1}^{2} \varphi_{2k}\, \rho(3-k)}{\rho(0) - \sum_{k=1}^{2} \varphi_{2k} \cdot \rho(k)} =$$

$$= \frac{\rho(3) - [\varphi_{21}\rho(2) - \varphi_{22}\rho(1)]}{1 - [\varphi_{21}\rho(1) - \varphi_{22}\rho(2)}$$

$$P_4^3 = P_3^2 (1 - \varphi_{33}^2)$$

$$\varphi_{32} = \varphi_{22} - \varphi_{33}\varphi_{21}$$

$$\varphi_{31} = \varphi_{21} - \varphi_{33} \cdot \varphi_{22}$$

m-Step Ahead Predictors

- Know how to find 1-step-ahead BLP X_{n+1}^{n}, but what about *m*-step-ahead BLP X_{n+m}^{n}?
  - Write $X_{n+m}^{n} = \varphi_{n1}^{(m)} X_{n} + \varphi_{n2}^{(m)} X_{n-1} + \cdots + \varphi_{nn}^{(m)} X_{1}$
  - To minimize MSE $\mathbb{E}[(X_{n+m} - X_{n+m}^{n})^{2}]$ must solve:

$$\mathbb{E}\left[\left(X_{n+m} - X_{n+m}^{n}\right) X_{n+1-k}\right] = 0 \Leftrightarrow$$

$$\Leftrightarrow \mathbb{E}\left[X_{n+m} X_{n+1-k}\right] - \sum_{j=1}^{n} \varphi_{nj}^{(m)} \mathbb{E}\left[X_{n+1-j} X_{n+1-k}\right] = 0 \Leftrightarrow$$

$$\Leftrightarrow \sum_{j=1}^{n} \varphi_{nj}^{(m)} \gamma(k-j) = \gamma(m+k-1), \quad \forall k = 1, \ldots, n$$

m-Step Ahead Predictors

- So, m-step-ahead BLP coefficients given by:

$$\sum_{j=1}^{n} \varphi_{nj}^{(m)} \gamma(k-j) = \gamma(m+k-1), \quad \forall k = 1, \ldots, n$$

$$\Leftrightarrow \quad \mathbf{\Gamma}_n \boldsymbol{\varphi}_n^{(m)} = \boldsymbol{\gamma}_n^{(m)} \;\Rightarrow\; \boldsymbol{\varphi}_n^{(m)} = \mathbf{\Gamma}_n^{-1} \boldsymbol{\gamma}_n^{(m)}$$

system can be solved with Durbin-Levinson

- where $\begin{cases} \boldsymbol{\varphi}_n^{(m)} = [\varphi_{n1}^{(m)} \quad \varphi_{n2}^{(m)} \quad \cdots \quad \varphi_{nn}^{(m)}]' \\ \boldsymbol{\gamma}_n^{(m)} = [\gamma(m) \quad \gamma(m+1) \quad \cdots \quad \gamma(m+n-1)]' \end{cases}$

- MSE P_{n+m}^{n} of m-step-ahead BLP is given by:

$$P_{n+m}^{n} = \mathbb{E}[(X_{n+m} - X_{n+m}^{n})^2] = \gamma(0) - \boldsymbol{\gamma}_n^{(m)\prime} \mathbf{\Gamma}_n^{-1} \boldsymbol{\gamma}_n^{(m)}$$

Innovations Algorithm

- To find coefficient of 1-step-ahead BLP X_{n+1}^{n} need to solve $\boldsymbol{\varphi}_n = \boldsymbol{\Gamma}_n^{-1}\boldsymbol{\gamma}_n$ (Durbin-Levinson)
  - For large n, this can be time-consuming
  - For, AR(p) model, however, solution is simple:
    - For $n \geq p: \ \varphi_{nj} = \varphi_j, \ 1 \leq j \leq p \ \ \& \ \ \varphi_{nj} = 0, \ \ p < j \leq n$
- It turns out there is another method to find X_{n+1}^{n} which gives simple solutions in the case of MA(q) models, called *Innovations Algorithm*

Innovations Algorithm

- Idea behind Innovations Algorithm: write X_{n+1}^{n} as linear combination of *previous innovations* (i.e. previous BLP prediction errors)
  - Define innovations as: $Z_t = X_t - X_t^{t-1}, \ \forall t \geq 1$
  - Rewrite BLP in terms of innovations as:

$$X_{n+1}^{n} = \sum_{j=1}^{n} \theta_{nj} Z_{n+1-j} = \sum_{j=1}^{n} \theta_{nj} (X_{n+1-j} - X_{n+1-j}^{n-j})$$

  - Then solve for θ_{nj}'s by minimizing MSE

Innovations Algorithm

- The BLP innovations' coefficients (θ_{nj}) can be found iteratively (in n) as follows:

Start with $X_1^0 = 0$ & $P_1^0 = \gamma(0)$, and for $n \geq 1$ set:

$$\theta_{n,n-k} = \frac{\gamma(n-k) - \sum_{j=0}^{k-1} \theta_{k,k-j}\theta_{n,n-j}P_{j+1}^j}{P_{k+1}^k}, \text{ for } k = 0,1,\ldots,n-1$$

$$X_{n+1}^n = \sum_{j=1}^n \theta_{nj} Z_{n+1-j} = \sum_{j=1}^n \theta_{nj}(X_{n+1-j} - X_{n+1-j}^{n-j})$$

$$P_{n+1}^n = \gamma(0) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 P_{j+1}^j$$

- Algorithm progresses as: $\boldsymbol{\theta}_1 \to \boldsymbol{\theta}_2 \to \boldsymbol{\theta}_3 \to \ldots$

Example

- For any MA(q) series, show that:

$$\theta_{n,n-k} = 0, \ \forall \, n > q \ \& \ n-k > q$$

For MA(q) $\Rightarrow \gamma(h) = 0, \ \forall \, h > q$

Let $n = q+1$ & $k = 0 \Rightarrow \Theta_{q+1,q+1} = \dfrac{\overbrace{\gamma(q+1)}^{=0} + \overbrace{\sum_{j=0}^{-1}}^{\text{empty sum} \, = 0}}{P_1^0} = \dfrac{0}{P_1^0} = 0$

$n = q+2$ & $\begin{cases} k=0 \Rightarrow \Theta_{q+2,q+2} = \dfrac{\overbrace{\gamma(q+2)}^{=0} + \cancel{\sum_{j=0}^{-1}}^{=0}}{P_1^0} = 0 \\ k=1 \Rightarrow \Theta_{q+2,q+1} = \dfrac{\gamma(q+1) + \sum_{j=0}^{0} \Theta_{k,k-j} \cdot \Theta_{n,n-j} \cdot P_{j+1}^{j}}{P_2^1} = \end{cases}$

$= \dfrac{0 + \Theta_{1,1} \cdot \overbrace{\Theta_{q+2,q+2}}^{0} \cdot P_1^0}{P_2^1} = 0$

& so on for $n \geq q+3$

Partial ACF

- For pure MA(q) model, $\rho(h)=0$ for $h>q$.
- For both AR(p) & ARMA(p,q) models, however, ρ(h) tails off to 0 as h increases
  - Can we distinguish AR from ARMA, as we can do with MA from ARMA based on ACF?
- *Partial ACF* (PACF) is another dependence measure that can expose differences between pure AR and ARMA models

Partial ACF

- Problem with ACF is that, even for simple AR(1) model, the correlation between X_t and its past carries over to infinite lags:
  - E.g. For $X_t=\varphi X_{t-1}+W_t$, we have

$$\gamma(2) = Cov(X_t, X_{t-2}) = Cov(\varphi X_{t-1} + W_t, X_{t-2}) =$$

$$= Cov(\varphi^2 X_{t-2} + \varphi W_{t-1} + W_t, X_{t-2}) = \varphi^2 \gamma(0)$$

- Ideally, would like to find the relation of X_t with X_{t+h}, after *removing* the effect of their in between variables $X_{t+1}, \ldots, X_{t+h-1}$

Partial ACF

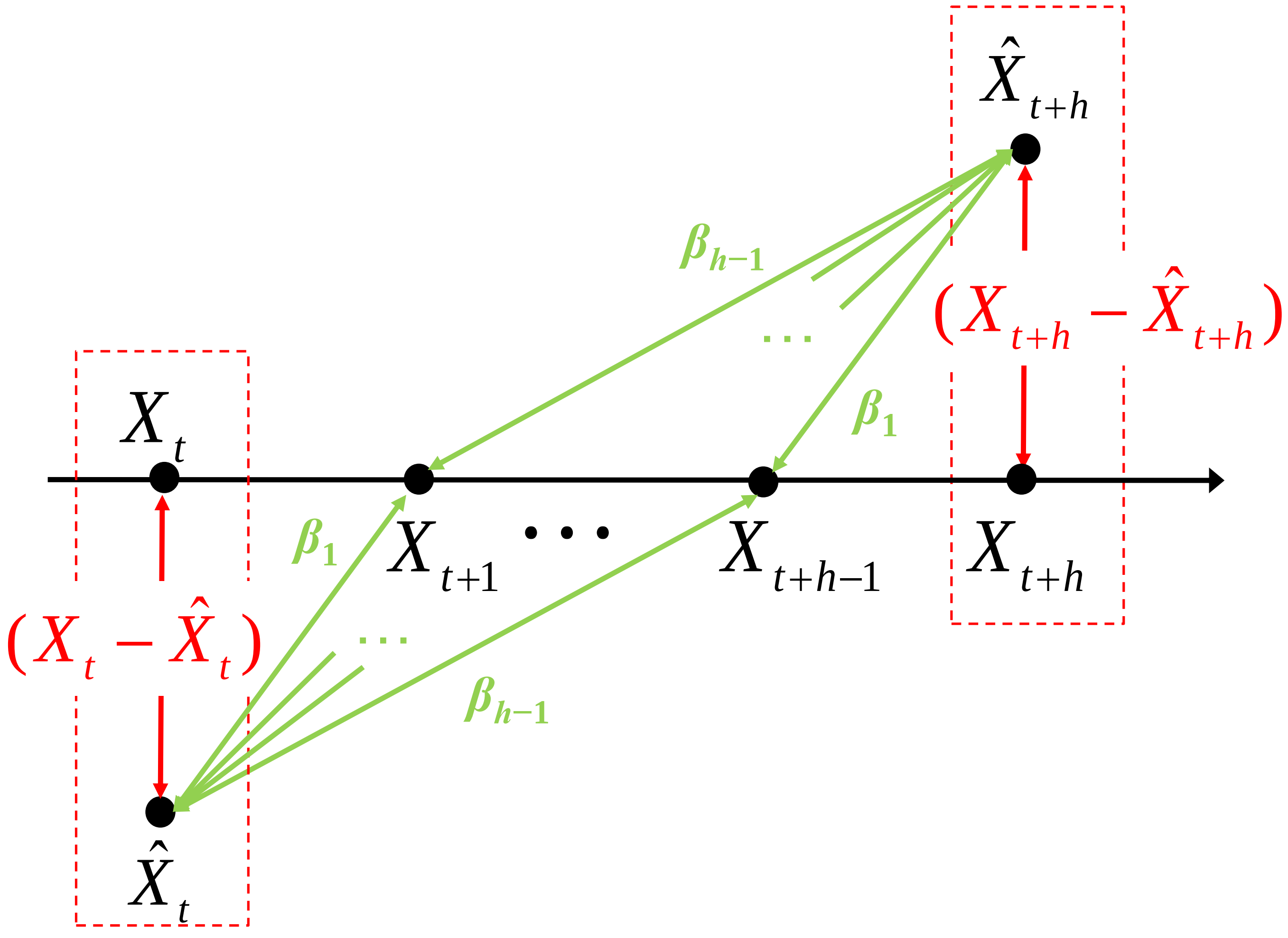
- For lag *h*, define the BLP's of X_{t+h} & X_t, both based on $\{X_{t+1},\ldots,X_{t+h-1}\}$

$$\hat{X}_{t+h} = \beta_1 X_{t+h-1} + \beta_2 X_{t+h-2} + \cdots + \beta_{h-1} X_{t+1}$$

$$\hat{X}_{t} = \beta_1 X_{t+1} + \beta_2 X_{t+2} + \cdots + \beta_{h-1} X_{t+h-1}$$

  - Predictors (i.e. β's) are *best* in that they *minimize* the Mean Squared Error (MSE) $\mathbb{E}[(X_{t+h} - \hat{X}_{t+h})^2]$
- To remove effect of intermediate variables, look at correlation of $(X_{t+h} - \hat{X}_{t+h})$ and $(X_t - \hat{X}_t)$

Partial ACF



$\hat{X}_{t+h}$
β_{h-1}
$\cdots$
$(X_{t+h} - \hat{X}_{t+h})$
X_t
β_1
β_1
X_{t+1}
$\cdots$
X_{t+h-1}
X_{t+h}
$(X_t - \hat{X}_t)$
$\cdots$
β_{h-1}
$\hat{X}_t$

Partial ACF

- The partial autocorrelation function (PACF) of a stationary time series $\{X_t\}$, denoted by φ_{hh} for h=1,2,…, is defined as:

$$\varphi_{11} = Cor(X_{t+1}, X_t) = \rho(1), \text{ and}$$

$$\varphi_{hh} = Cor[(X_{t+h} - \hat{X}_{t+h}),(X_t - \hat{X}_t)], \ \forall h \geq 2$$

  - We will see how to calculate the PACF later
  - For now, we will just look at its overall behavior for AR and MA models

Partial ACF

- For pure AR(p) model $X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + W_t$ the PACF has the following properties:

$$\begin{cases} \varphi_{hh} \text{ not necessarily } 0, \text{ for } 1 \le h < p \\ \varphi_{hh} = \varphi_{pp} = \varphi_p, \text{ for } h = p \\ \varphi_{hh} = 0, \text{ for } h > p \end{cases}$$

- For MA / ARMA model, the PACF is not necessarily 0, but tails off to 0 as h increases
  - Heuristic proof: write MA / ARMA as AR(∞) from invertibility, and use properties above

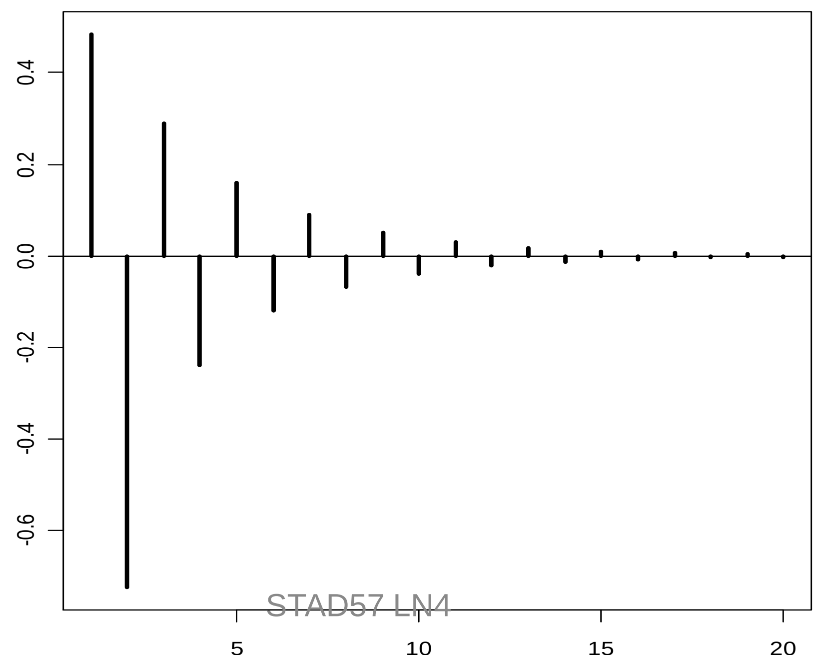
Finding PACF in R

- R function `ARMAacf()` can also calculate φ_{hh}
  - E.g. for $X_t = .7X_{t-1} - .3X_{t-2} + W_t + .5W_{t-1} - .2W_{t-2}$

list of φ's — list of θ's — # of φ_{hh}'s — option for PACF

```
ARMAacf(ar=c(.7,-.3), ma=c(.5,-.2), 20, pacf=TRUE)
```

Resulting PACF:
(first 20 lags)



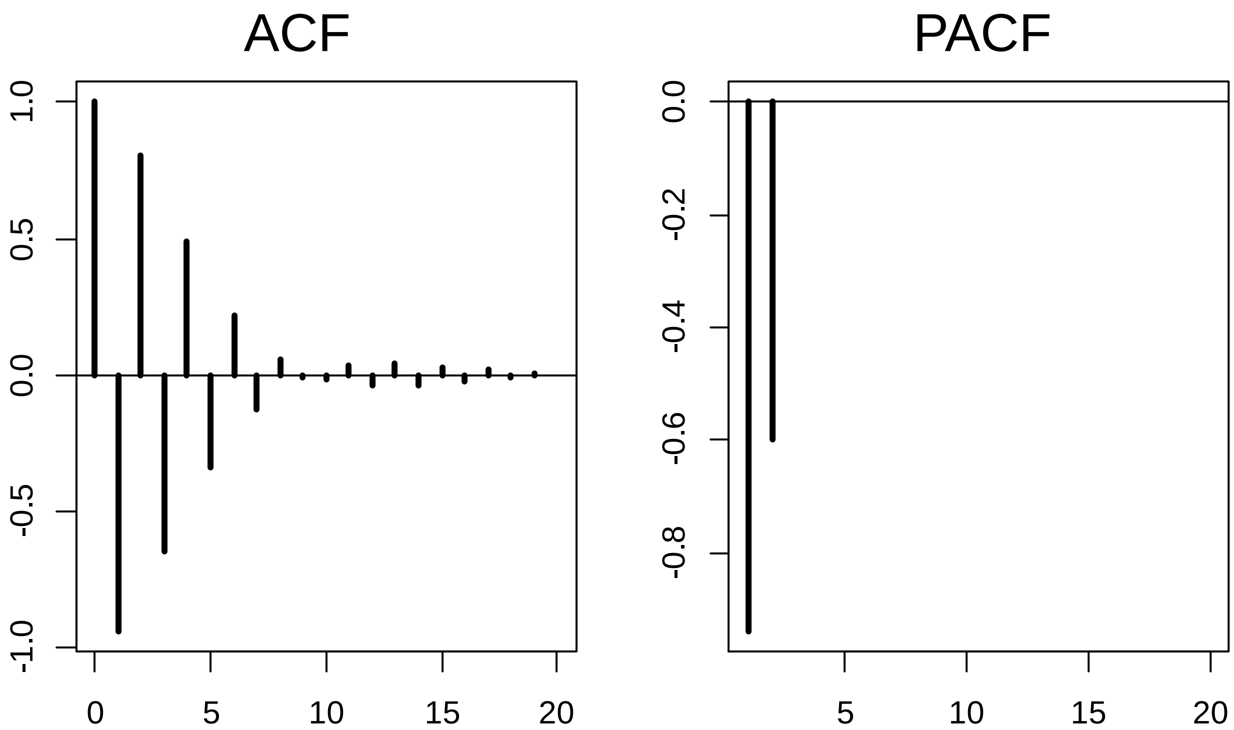
Model Identification

- Can use properties of (sample) ACF & PACF to identify the *order* of an ARMA model
- The following table describes their behavior

	AR(p)	**MA(q)**	**ARMA(p,q)**
ACF	Tails off	Cuts off (for h>q)	Tails off
PACF	Cuts off (for h>p)	Tails off	Tails off

Example

- Try to identify order of ARMA model with:

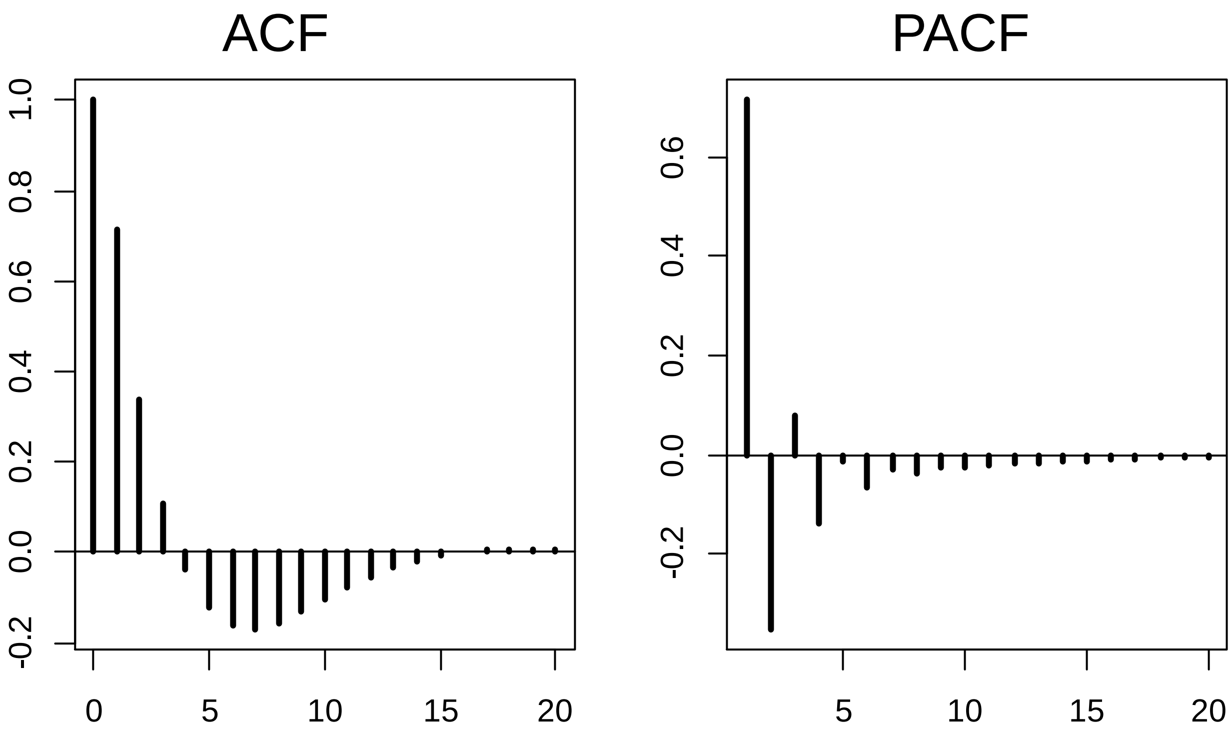


PACF cuts off at lag > 2

=> AR(p=2)

Example

- Try to identify order of ARMA model with:

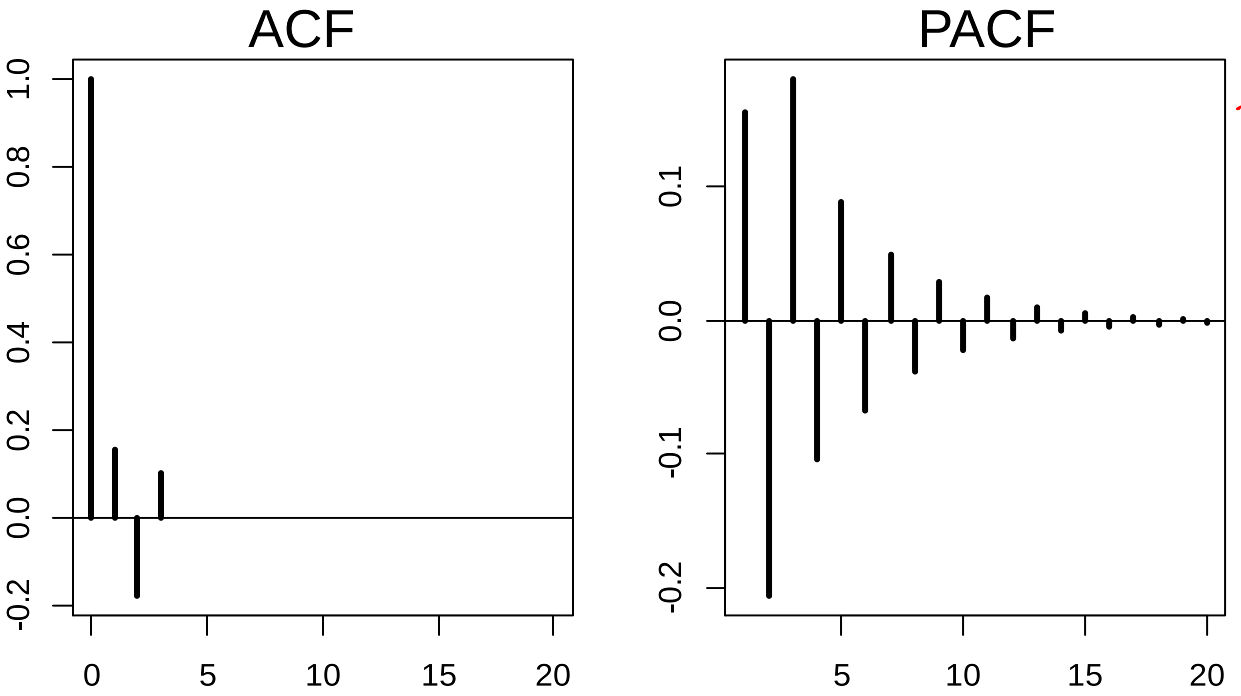


both ACF & PACF tail off

→ use general ARMA(p,q)

Example

- Try to identify order of ARMA model with:



→ ACF cuts off

⇒ MA(q=3)

PACF

- Have defined PACF of a TS to be given by:

$$\varphi_{11} = Cor(X_{t+1}, X_t) = \rho(1), \text{ \&}$$

$$\varphi_{hh} = Cor[(X_{t+h} - \hat{X}_{t+h}), (X_t - \hat{X}_t)], \ \forall h \geq 2$$

where: $\begin{cases} \hat{X}_{t+h} = \beta_1 X_{t+h-1} + \beta_2 X_{t+h-2} + \cdots + \beta_{h-1} X_{t+1} \\ \hat{X}_t = \beta_1 X_{t+1} + \beta_2 X_{t+2} + \cdots + \beta_{h-1} X_{t+h-1} \end{cases}$

are the BLP's of X_{t+h} & X_t, based on $\{X_{t+1}, \ldots, X_{t+h-1}\}$

- Now look at relationship of PACF to $\boldsymbol{\varphi}$'s, and how to calculate it

PACF

- First, note that $\hat{X}_{t+h}$ is 1-step-ahead BLP of X_{t+h} based on $\mathbf{X} = \left[X_{t+h-1} \ X_{t+h-2} \ \cdots \ X_{t+1} \right]'$

$$\Rightarrow \hat{X}_{t+h} = \varphi_{h-1,1} X_{t+h-1} + \varphi_{h-1,2} X_{t+h-2} + \cdots + \varphi_{h-1,h-1} X_{t+1}$$

  - Since the optimal coefficients are the same for any t (by stationarity), assume $t=0$ for simplicity

i.e. $\begin{cases} \varphi_{11} = Cor(X_1, X_0) = \rho(1), \ \& \\ \varphi_{hh} = Cor[(X_h - \hat{X}_h),(X_0 - \hat{X}_0)], \ \forall h \geq 2 \end{cases}$, and

$$\hat{X}_h = X_h^{h-1} = \varphi_{h-1,1} X_{h-1} + \cdots + \varphi_{h-1,h-1} X_1, \ \forall h \geq 2$$

PACF

- Also note that $\hat{X}_0$, the BLP of X_0 based on $\mathbf{X} = \left[X_{h-1} \; X_{h-2} \cdots X_1 \right]'$ is given by

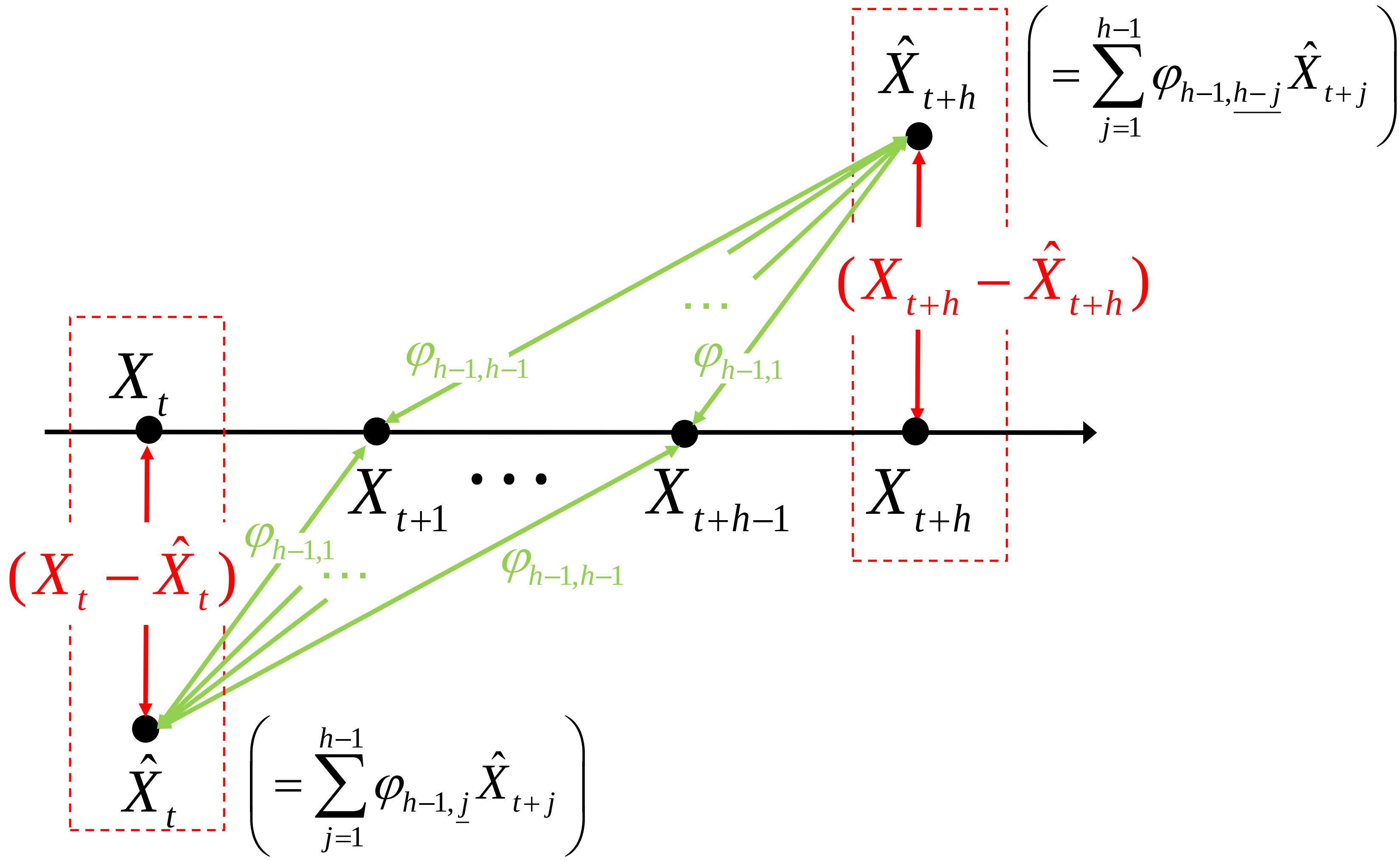
$$\hat{X}_0 = \varphi_{h-1,h-1} X_{h-1} + \varphi_{h-1,h-2} X_{h-2} + \cdots + \varphi_{h-1,1} X_1, \;\; \forall h \geq 2$$

  - Uses the *same coefficients* as $\hat{X}_h$ in *reverse order*
  - Easy to prove, since BLP of X_0 solves system:

$$\mathbb{E}\left[\left(X_0 - \hat{X}_0 \right) X_k \right] = 0 \Leftrightarrow \sum_{j=1}^{h-1} \varphi_{h-1,j} \gamma(k-j) = \gamma(k), \;\; \forall k = 1, \ldots, h-1$$

which is the same as for BLP of $\hat{X}_h$

PACF


$\hat{X}_{t+h}$
$\left(=\sum_{j=1}^{h-1}\varphi_{h-1,\underline{h-j}}\hat{X}_{t+j}\right)$
$(X_{t+h}-\hat{X}_{t+h})$
$\varphi_{h-1,h-1}$
$\varphi_{h-1,1}$
X_t
X_{t+1}
X_{t+h-1}
X_{t+h}
$(X_t-\hat{X}_t)$
$\varphi_{h-1,1}$
$\varphi_{h-1,h-1}$
$\hat{X}_t$
$\left(=\sum_{j=1}^{h-1}\varphi_{h-1,\underline{j}}\hat{X}_{t+j}\right)$

PACF

- Thus, we can write $\hat{X}_h = \boldsymbol{\varphi}'_{h-1}\mathbf{X}$ & $\hat{X}_0 = \breve{\boldsymbol{\varphi}}'_{h-1}\mathbf{X}$ where: $\mathbf{X} = \left[X_{h-1} \; X_{h-2} \cdots X_1 \right]'$, $\boldsymbol{\varphi}_{h-1} = \left[\varphi_{h-1,1} \; \varphi_{h-1,2} \cdots \varphi_{h-1,h-1} \right]'$ & $\breve{\boldsymbol{\varphi}}_{h-1} = \left[\varphi_{h-1,h-1} \; \varphi_{h-1,h-2} \cdots \varphi_{h-1,1} \right]'$, i.e. the reverse of $\boldsymbol{\varphi}_{h-1}$
- We can now show that $\varphi_{hh} = \mathrm{Cor}[(X_{t+h} - \hat{X}_{t+h}), (X_t - \hat{X}_t)], \; \forall h \geq 2$ where φ_{hh} is given by BLP of X_{h+1}^{h}
  - Use fact that $\mathrm{Cov}\left(\mathbf{a}'\mathbf{X}, \mathbf{b}'\mathbf{Y}\right) = \mathbf{a}'\,\mathrm{Cov}(\mathbf{X},\mathbf{Y})\mathbf{b}$, where $\mathbf{a},\mathbf{b}$ are constant vectors, $\mathbf{X},\mathbf{Y}$ are random vectors, and $\mathrm{Cov}(\mathbf{X},\mathbf{Y})$ is covariance matrix of $\mathbf{X},\mathbf{Y}$

$$\operatorname{Cor}(X_h - \hat{X}_h, X_0 - \hat{X}_0) = \frac{\operatorname{Cov}(X_h - \hat{X}_h, X_0 - \hat{X}_0)}{\sqrt{\operatorname{Var}(X_h - \hat{X}_h) \cdot \operatorname{Var}(X_0 - \hat{X}_0)}}$$

which by stationarity, can show $\operatorname{Var}(X_h - \hat{X}_h) = \operatorname{Var}(X_0 - \hat{X}_0) \Rightarrow$

$$\Rightarrow \operatorname{Cor}(X_h - \hat{X}_h, X_0 - \hat{X}_0) = \frac{\operatorname{Cov}(X_h - \hat{X}_h, X_0 - \hat{X}_0)}{\operatorname{Var}(X_h - \hat{X}_h)}$$

<u>For denominator</u>: $\operatorname{Var}(X_h - \hat{X}_h) = \operatorname{Var}(X_h - \phi_{h-1}' \underline{X}) =$

$$= \operatorname{Var}(X_h) + \operatorname{Var}(\phi_{h-1}' \cdot \underline{X}) - 2 \cdot \operatorname{Cov}(X_h, \phi_{h-1}' \underline{X}) =$$

$$= \gamma(0) + \phi_{h-1}' \cdot \underbrace{\operatorname{Cov}(\underline{X}, \underline{X})}_{= \Gamma_{h-1}} \cdot \phi_{h-1} - 2 \cdot \phi_{h-1}' \cdot \underbrace{\operatorname{Cov}(X_h, \underline{X})}_{= \gamma_{h-1}} =$$

$$= \gamma(0) + \phi_{h-1}' \Gamma_{h-1} \cdot \phi_{h-1} - 2 \phi_{h-1}' \cdot \gamma_{h-1} = \qquad (\phi_{h-1} = \Gamma_{h-1}^{-1} \cdot \gamma_{h-1})$$

$$= \gamma(0) + \phi_{h-1}' \cancel{\Gamma_{h-1} \cdot \Gamma_{h-1}^{-1}} \gamma_{h-1} - \cancel{2} \phi_{h-1}' \gamma_{h-1} =$$

$$= \gamma(0) - \phi_{h-1}' \cdot \gamma_{h-1}$$

For numerator: $\mathrm{Cov}(X_h - \hat{X}_h, X_0 - \hat{X}_0) = \mathrm{Cov}(X_h - \varphi_{h-1}' \underline{X}, X_0 - \breve{\varphi}_{h-1}' \underline{X})$

$$= \mathrm{Cov}(X_h, X_0) - \varphi_{h-1}' \mathrm{Cov}(\underline{X}, X_0) - \breve{\varphi}_{h-1}' \mathrm{Cov}(X_h, \underline{X}) + \varphi_{h-1}' \mathrm{Cov}(\underline{X}, \underline{X}) \breve{\varphi}_{h-1} =$$

$$= \gamma(h) - \underbrace{\varphi_{h-1}' \cdot \breve{\gamma}_{h-1}}_{\text{equal}} - \underbrace{\breve{\varphi}_{h-1}' \cdot \gamma_{h-1}}_{\text{equal}} + \underbrace{\varphi_{h-1}' \cdot \cancel{\Gamma_{h-1}}}_{= \gamma_{h-1}' \cancel{\Gamma_{h-1}^{-1}}} \cdot \breve{\varphi}_{h-1} =$$

$$= \gamma(h) - 2 \cdot \underbrace{\varphi_{h-1}' \breve{\gamma}_{h-1}}_{\text{equal}} + \cancel{\underbrace{\gamma_{h-1}' \cdot \breve{\varphi}_{h-1}}}$$

(inner products for reverse vectors equal to inner products for original vectors)

$$\Rightarrow \mathrm{Cov}(X_h - \hat{X}_h, X_0 - \hat{X}_0) = \gamma(h) - \varphi_{h-1}' \cdot \breve{\gamma}_{h-1}$$

$$\Rightarrow \text{PACF @ lag } h: \quad \varphi_{hh} = \frac{\gamma(h) - \varphi_{h-1}' \breve{\gamma}_{h-1}}{\gamma(0) - \varphi_{h-1}' \gamma_{h-1}} =$$

(expand inner products $\varphi' \breve{\gamma}$ & $\varphi \gamma$)

$$\left\{\begin{array}{l} \varphi_{h-1}' \breve{\gamma}_{h-1} = [\varphi_{h-1,1} \; \cdots \; \varphi_{h-1,h-1}] \cdot \begin{bmatrix} \gamma(h-1) \\ \vdots \\ \gamma(1) \end{bmatrix} = \sum_{k=1}^{h-1} \varphi_{h-1,k} \cdot \gamma(h-k) \\ \varphi_{h-1}' \gamma_{h-1} = [\varphi_{h-1,1} \; \cdots \; \varphi_{h-1,h-1}] \cdot \begin{bmatrix} \gamma(1) \\ \vdots \\ \gamma(h-1) \end{bmatrix} = \sum_{k=1}^{h-1} \varphi_{h-1,k} \cdot \gamma(k) \end{array}\right)$$

$$\Rightarrow \varphi_{hh} = \frac{\gamma(h) - \sum_{k=1}^{h-1} \varphi_{h-1,k} \cdot \gamma(h-k)}{\gamma(0) - \sum_{k=1}^{h-1} \varphi_{h-1,k} \cdot \gamma(k)},$$

which is exactly the D-L algorithm formula for $\varphi_{h,h}$

PACF

- So, given autocovariance γ(h) / ACF ρ(h), we can find PACF φ_{hh} iteratively, using the Durbin-Levinson algorithm: $\varphi_{11} = \rho(1)$ & for $h \geq 2$

$$\varphi_{hh} = \frac{\gamma(h) - \sum_{k=1}^{h-1} \varphi_{h-1,k}\gamma(h-k)}{\gamma(0) - \sum_{k=1}^{h-1} \varphi_{h-1,k}\gamma(k)} = \frac{\rho(h) - \sum_{k=1}^{h-1} \varphi_{h-1,k}\rho(h-k)}{1 - \sum_{k=1}^{h-1} \varphi_{h-1,k}\rho(k)}$$

$$\varphi_{hk} = \varphi_{h-1,k} - \varphi_{hh}\varphi_{h-1,h-k}, \quad k = 1, 2, \ldots, h-1 \text{ (if } h \geq 2)$$

- To get the *sample* PACF, just use algorithm with sample autocovariance / sample ACF

Example

For AR(2) we know: $\begin{cases} \rho(0) = 1 \\ \rho(1) = \frac{\varphi_1}{1-\varphi_2} \\ \rho(2) = \frac{\varphi_1^2}{1-\varphi_2}, \cdots \rho(h) = \frac{\varphi_1^h}{1-\varphi_2} \quad \forall h \geq 2 \end{cases}$

- Find PACF of AR(2): $X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + W_t$

For $h=1$: $\varphi_{11} = \frac{\rho(1)}{\cancel{\rho(0)}=1} = \rho(1) = \frac{\varphi_1}{1-\varphi_2}$

For $h=2$: $\varphi_{2,2} = \frac{\rho(2) - \varphi_{11}\cdot\rho(1)}{\rho(0) - \varphi_{11}\cdot\rho(1)} = \cdots = \varphi_2$

For $h \geq 3$: $\varphi_{3,3} = 0$

$\varphi_{4,4} = 0$

$\vdots$

$= \frac{\rho(3) - \overset{\varphi_1}{\varphi_{21}}\rho(2) - \overset{\varphi_2}{\varphi_{22}}\rho(1)}{\rho(0) - \varphi_{21}\rho(1) - \varphi_{22}\rho(2)} = \frac{0}{\cdots} = 0$

b/c for $h \geq p$ in AR(p) model, $\varphi_{h,j} = \varphi_j$, $j = 1, \ldots, p$; $\varphi_{h,j} = 0$, $j > p$

from recurrence eqn for ACF of AR(p) model

Forecasting ARMA Processes

- Consider now general (causal & invertible) ARMA(p,q) model: $\varphi(B)X_t = \theta(B)W$
  - For large n, neither Durbin-Levinson nor Innovations algorithms are effective
- Can build (approximate) forecasts & examine their behavior by assuming we know *infinite series past*: $\mathbf{X}_{n-} = \{X_n, X_{n-1}, \ldots, X_1, X_0, X_{-1}, \ldots\}$
  - Denote minimum MSE m-step-ahead forecast by

$$\tilde{X}_{n+m} = \mathbb{E}\left[X_{n+m} \mid \mathbf{X}_{n-}\right] = \mathbb{E}\left[X_{n+m} \mid X_n, X_{n-1}, \ldots\right]$$

Forecasting ARMA Processes

- Given infinite past $\mathbf{X}_{n-}$, we have:

$$\tilde{X}_{n+m} = -\sum_{j=1}^{m-1} \pi_j \tilde{X}_{n+m-j} - \sum_{j=m}^{\infty} \pi_j X_{n+m-j} = \sum_{j=m}^{\infty} \psi_j W_{n+m-j}$$

$$P_{n+m}^{n} = \mathbb{E}\left[\left(X_{n+m} - \tilde{X}_{n+m}\right)^2\right] = \sigma_w^2 \sum_{j=0}^{m-1} \psi_j^2$$

where π_j, ψ_j are the causal / invertible weights

- Proof:

From causality: $X_{n+m} = \sum_{j=0}^{\infty} \psi_j W_{n+m-j}$ ①

From invertibility: $W_{n+m} = \sum_{j=0}^{\infty} \pi_j X_{n+m-j}$ ②

Take conditional expectations of (1) given infinite past $\underline{X}_{n-} = \{X_n, X_{n-1}, \dots, X_1, X_0, X_{-1}, X_{-2} \dots\}$

$$\Rightarrow \underbrace{\mathbb{E}[X_{n+m} \mid \underline{X}_{n-}]}_{\tilde{X}_{n+m}} = \sum_{j=0}^{\infty} \psi_j \cdot \underbrace{\mathbb{E}[W_{n+m-j} \mid \underline{X}_{n-}]}_{\text{call this } \tilde{W}_{n+m-j}} \Rightarrow \tilde{X}_{n+m} = \sum_{j=0}^{\infty} \psi_j \tilde{W}_{n+m-j}$$

$$\tilde{W}_{n+m-j} = \mathbb{E}[W_{n+m-j} \mid \underline{X}_{n-}] = \begin{cases} 0, & \text{if } n+m-j > n \Rightarrow j < m \\ W_{n+m-j}, & \text{if } n-m-j \le n \Rightarrow j \ge m \end{cases}$$

$$\Rightarrow \tilde{X}_{n+m} = \sum_{j=0}^{m-1} \psi_j \cdot 0 + \sum_{j=m}^{\infty} \psi_j \cdot W_{n+m-j} = \sum_{j=m}^{\infty} \psi_j W_{n+m-j}$$

Take conditional expectation of (2) given $\underline{X}_{n-}$

$$\Rightarrow \mathbb{E}[W_{n+m} \mid \underline{X}_{n-}] = \sum_{j=0}^{\infty} \Pi_j \mathbb{E}[X_{n+m-j} \mid \underline{X}_{n-}] \Rightarrow$$

$$0 = \underbrace{\Pi_0}_{=1} \underbrace{\mathbb{E}[X_{n+m} \mid \underline{X}_{n-}]}_{=\tilde{X}_{n+m}} + \sum_{j=1}^{m-1} \Pi_j \underbrace{\mathbb{E}[X_{n+m-j} \mid \underline{X}_{n-}]}_{\tilde{X}_{n+m-j}} + \sum_{j=m}^{\infty} \Pi_j \overbrace{\mathbb{E}[X_{n+m-j} \mid X_{n-}]}^{X_{n+m-j}}$$

$$\Rightarrow 0 = \tilde{X}_{n+m} + \sum_{j=1}^{m-1} \pi_j \tilde{X}_{n+m-j} + \sum_{j=m}^{\infty} \pi_j X_{n+m-j} \Rightarrow$$

$$\Rightarrow \tilde{X}_{n+m} = -\sum_{j=1}^{m-1} \pi_j \tilde{X}_{n+m-j} - \sum_{j=m}^{\infty} \pi_j X_{n+m-j}$$

From $(X_{n+m} - \tilde{X}_{n+m}) = \left(\sum_{j=0}^{\infty} \psi_j W_{n+m-j} - \sum_{j=m}^{\infty} \psi_j W_{n+m-j}\right) =$

$$= \sum_{j=0}^{m-1} \psi_j W_{n+m-j} \Rightarrow$$

$$\Rightarrow P_{n+m}^{n} = \mathbb{E}\left[(X_{n+m} - \tilde{X}_{n+m})^2\right] = \mathbb{E}\left[\left(\sum_{j=0}^{m-1} \psi_j W_{n+m-j}\right)^2\right] =$$

$$= \sum_{j=0}^{m-1} \psi_j^2 \cdot \underbrace{\mathbb{E}\left[W_{n+m-j}^2\right]}_{=\sigma_w^2} = \sigma_w^2 \cdot \sum_{j=0}^{m-1} \psi_j^2$$

Long-Range Forecasts

- From representation of MMSE m-step-ahead forecast (for non-zero-mean process):

$$\tilde{X}_{n+m} = \mu + \sum_{j=m}^{\infty} \psi_j W_{n+m-j}$$

as m→∞, we have:

- $\tilde{X}_{n+m} \to \mu$ (since $\sum_{j=0}^{\infty} |\psi_j| < \infty$ for ψ weights $\Rightarrow$

  $\Rightarrow$ tail sums $\sum_{j=m}^{\infty} \psi_j \to 0$ as $m \to \infty$)

- $P_{n+m}^{n} \to \sigma_w^2 \sum_{j=0}^{\infty} \psi_j^2 = \gamma(0)$

  (MSE increases to variance $\gamma(0)$ exponentially fast)

Truncated ARMA Prediction

$X_0 = 0, \; X_{-1} = 0$

- In practice, cannot know infinite past $\mathbf{X}_{n-}$, but we can use finite past $\{X_1, \ldots, X_n\}$ to make truncated predictions, denoted by $\tilde{X}_{n+m}^{n}$
  - From $\tilde{X}_{n+m} = -\sum_{j=1}^{m-1} \pi_j \tilde{X}_{n+m-j} - \sum_{j=m}^{\infty} \pi_j X_{n+m-j}$

    set $X_t = 0, \; \forall t \le 0 \Rightarrow \sum_{j=n+m}^{\infty} \pi_j X_{n+m-j} = 0$
  - Truncated prediction becomes:

$$\tilde{X}_{n+m}^{n} = -\sum_{j=1}^{m-1} \pi_j \tilde{X}_{n+m-j}^{n} - \sum_{j=m}^{n+m-1} \pi_j X_{n+m-j}$$

Truncated ARMA Prediction

- For ARMA(p,q) model $\varphi(B)X_t = \theta(B)W_t$, truncated predictions can be recast as:

$$\tilde{X}^n_{n+m} = \varphi_1 \tilde{X}^n_{n+m-1} + \cdots + \varphi_p \tilde{X}^n_{n+m-p} + \theta_1 \tilde{W}^n_{n+m-1} + \cdots + \theta_q \tilde{W}^n_{n+m-q}$$

- where: $\begin{cases} \tilde{X}^n_t = 0, \forall t \le 0 \ \ \& \ \ \tilde{X}^n_t = X_t, \forall 1 \le t \le n \\ \tilde{W}^n_t = 0, \forall\, t \le 0 \text{ or } t > n, \ \ \& \\ \tilde{W}^n_t = \varphi(B)\tilde{X}^n_t - \theta_1 \tilde{W}^n_{t-1} \cdots - \theta_q \tilde{W}^n_{t-q}, \forall 1 \le t \le n \end{cases}$

- Prediction MSE approximated by $P^n_{n+m} = \sigma^2_w \sum_{j=0}^{m-1} \psi^2_j$

Example

- Consider ARMA(1,1) model: $X_t = \underbrace{.9}_{\phi_1} X_{t-1} + W_t - \underbrace{.5}_{\theta_1} W_{t-1}$
  - If X_1=1.4, X_2=.5, & X_3=−.3, find 1- to 3- step-ahead truncated forecasts & their MSE

$\tilde{X}^n_{n+m} = .9\tilde{X}_{t-1} - .5\tilde{W}_{t-1}$ $\left\{ \begin{array}{l} n=3 \\ m=1,2,3 \end{array} \right.$

where $\left\{ \begin{array}{l} \tilde{X}^3_t = X_t \ \forall\ t=1,2,3 \ \& \ \tilde{X}^3_t = 0,\ \forall\ t \leq 0 \\ \tilde{W}^3_t = 0,\ \forall\ t \leq 0 \ \& \ t \geq 4 \\ \tilde{W}^3_t = \varphi(B) \cdot \tilde{X}^3_t - \theta_1 \tilde{W}^3_{t-1} = \tilde{X}^3_t - .9\tilde{X}^3_{t-1} + .5\tilde{W}^3_{t-1},\ \forall t=1,2,3 \end{array} \right.$

$\Rightarrow \left\{ \begin{array}{l} \tilde{W}^3_1 = \tilde{X}^3_1 - .9\cancel{\tilde{X}^3_0}^{=0} + .5\cancel{\tilde{W}^3_0}^{=0} = \tilde{X}^3_1 = 1.4 \\ \tilde{W}^3_2 = \tilde{X}^3_2 - .9\tilde{X}^3_1 + .5\tilde{W}^3_1 = .5 - .9 \times 1.4 + .5 \times 1.4 = -.06 \\ \tilde{W}^3_3 = \tilde{X}_3 - .9\tilde{X}^3_2 + .5\tilde{W}^3_2 = -.3 - .9 \times .5 + .5 \times (-.06) = -.78 \end{array} \right.$

$\underline{\underline{m=1}}$ $$\tilde{X}_4^3 = .9\,\tilde{X}_3^3 - .5\,\tilde{W}_3^3 = .9\times(-.3) - .5(-.78) = .12$$

$\underline{\underline{m=2}}$ $$\tilde{X}_5^3 = .9\,\tilde{X}_4^3 - .5\,\cancel{\tilde{W}_4^3}^{\,0} = .9(.12) = .108$$

$\underline{\underline{m=3}}$ $$\tilde{X}_6^3 = .9\,\tilde{X}_5^3 - .5\,\cancel{\tilde{W}_5^3}^{\,=0} = .9(.108) = .0972$$

For MSE P^n_{n+m}, we know for ARMA(1,1), ψ-weights are: $\psi_j = (\varphi+\theta)\varphi^{j-1},\ \forall j \geq 1 \Rightarrow \psi_j = (.9-.5)\times(.9)^{j-1}$
$= .4\times(.9)^{j-1}$

Also $$P^n_{n+m} \approx \sigma_w^2 \cdot \sum_{j=0}^{m-1} \psi_j^2 = \sigma_w^2 \cdot \left[1 + (.4)^2 \sum_{j=1}^{m-1} (.9^2)^{j-1}\right]$$

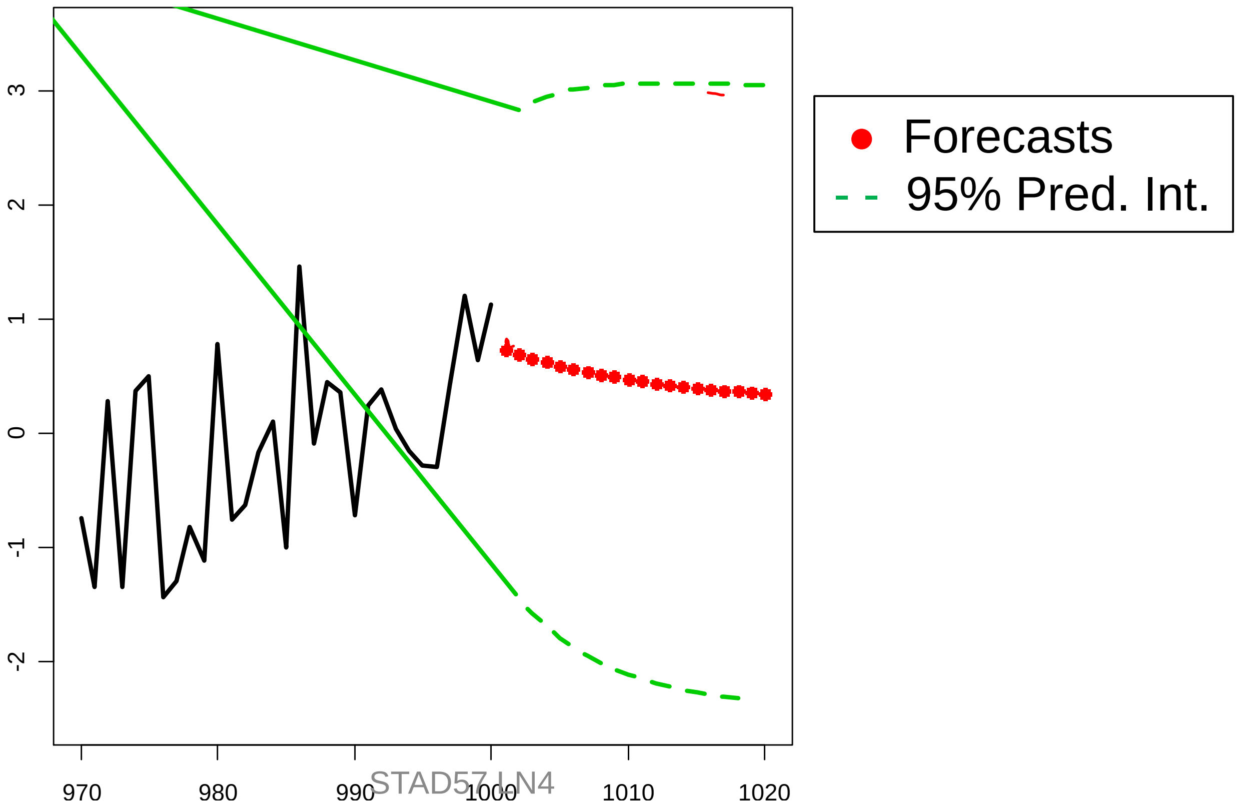
$$\Rightarrow \begin{cases} P_4^3 = \sigma_w^2 \\ P_5^3 = \sigma_w^2\left(1+(.4)^2\right) = \sigma_w^2 \times 1.16 \\ P_6^3 = \sigma_w^2 \cdot \left(1+(.4)^2 + (.4)^2\cdot(.9)^2\right) = \sigma_w^2 \cdot 1.2896 \end{cases}$$

Prediction Intervals

- Given forecast X_{n+m}^{n} (either exact / truncated) and its MSE P_{n+m}^{n}, can build (1−α) prediction intervals as: $X_{n+m}^{n} \pm c_{\alpha/2}\sqrt{P_{n+m}^{n}}$
  - Where $c_{\alpha/2}$ is $\alpha/2$ critical value from appropriate distribution (usually Normal)
    - E.g. For Normal 95% prediction interval → $c_{2.5\%} = 1.96$
  - For simultaneous prediction intervals, can adjust significance level using Bonferroni correction

Example

- Prediction intervals for 1- to 20-step-ahead forecasts of ARMA(1,1) model $X_t = .9X_{t-1} + W_t - .5W_{t-1}$



ARMA Forecasting

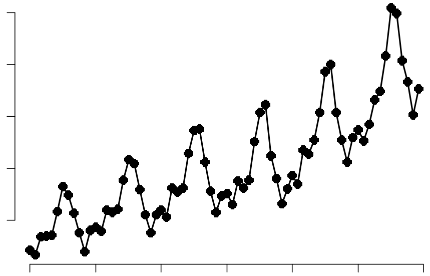
Overall, we have:

- For small n, use exact forecasts
  - Durbin-Levinson algorithm
    - Well suited for pure AR(p) models
    - Also used to find PACF
  - Innovations algorithm
    - Well suited for pure MA(q) models

$O(n^2)$

- For large n, use truncated forecasts

$O(n)$



5. ARMA Estimation

STAD57 F19

Sotirios Damouras

ARMA Estimation

- Given data $\{X_1,\ldots,X_n\}$ from causal / invertible ARMA(p,q) model, want to *estimate parameters* $\varphi_1,\ldots,\varphi_p,\theta_1,\ldots,\theta_q,\sigma_w^2$
  - For now, assume model *order* (p,q) is *known*
    - Will look at model selection later
  - Assume $\{X_t\}$ is zero-mean; if not, de-mean
  - Look at 3 types of estimation:
    - Method of Moments (MM)
    - Maximum Likelihood (ML)
    - Least Squares (LS)

ARMA Estimation

- Estimation is easier for pure AR(p) models

$$X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + W_t$$

  - Can use MM, ML, or LS estimation, with usually *closed form* solutions
- Estimation is more involved for general ARMA(p,q) models

$$X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + W_t + \theta_1 W_{t-1} + \cdots + \theta_q W_{t-q}$$

  - Can only use LS or ML estimation, with *numerical* solutions

Yule-Walker AR(p) Estimation

- For AR(p) model $X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + W_t$ already know $\gamma(h) = \varphi_1 \gamma(h-1) + \cdots + \varphi_p \gamma(h-p), \forall h \ge 1$ But we *also* have $\sigma_w^2 = \gamma(0) - \varphi_1 \gamma(1) - \cdots - \varphi_p \gamma(p)$
  - Proof: $Var[X_t] = \gamma(0) = Var[\varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + W_t] \Rightarrow$

$\Rightarrow \gamma(0) = Var[\varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p}] + Var[W_t] \Rightarrow$

$\Rightarrow \gamma(0) = Cov(\varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p}, \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p}) + \sigma_w^2$

$= \varphi_1 \cdot [\varphi_1 Cov(X_{t-1}, X_{t-1}) + \cdots + \varphi_p \cdot Cov(X_{t-1}, X_{t-p})$
$+ \vdots$
$+ \varphi_p \cdot [\varphi_1 Cov(X_{t-p}, X_{t-1}) + \cdots + \varphi_p Cov(X_{t-p}, X_{t-p}) + \sigma_w^2 =$

$\Rightarrow \gamma(0) = \varphi_1 \cdot (\underbrace{\varphi_1 \cdot \gamma(0) + \cdots + \varphi_p \gamma(p-1)}_{= \gamma(1)}) + \cdots + \varphi_p \cdot [\underbrace{\varphi_1 \cdot \gamma(p-1) + \cdots + \varphi_p \gamma(0)}_{\gamma(p)}) + \sigma_w^2 =$

Yule-Walker AR(p) Estimation

- Yule-Walker equations:

$$\begin{cases} \gamma(h) = \varphi_1 \gamma(h-1) + \cdots + \varphi_p \gamma(h-p),\ h = 1, \ldots, p \\ \sigma_w^2 = \gamma(0) - \varphi_1 \gamma(1) - \cdots - \varphi_p \gamma(p) \end{cases} \Leftrightarrow$$

$$\Leftrightarrow \begin{cases} \gamma(1) = \varphi_1 \gamma(0) + \varphi_2 \gamma(1) + \cdots + \varphi_p \gamma(p-1) \\ \qquad \vdots \\ \gamma(p) = \varphi_1 \gamma(p-1) + \varphi_2 \gamma(p-1) + \cdots + \varphi_p \gamma(0) \\ \sigma_w^2 = \gamma(0) - \varphi_1 \gamma(1) - \cdots - \varphi_p \gamma(p) \end{cases}$$

$\Leftrightarrow \mathbf{\Gamma}_p \boldsymbol{\varphi} = \boldsymbol{\gamma}_p,\ \ \sigma_w^2 = \gamma(0) - \boldsymbol{\varphi}' \boldsymbol{\gamma}_p,$ where $\boldsymbol{\varphi}' = \left[\varphi_1\ \varphi_2\ \cdots\ \varphi_p \right]$

Yule-Walker AR(p) Estimation

- MM estimation: replace 2nd order moments $\gamma(h)$ with sample estimates $\hat{\gamma}(h)$ & solve for $(\hat{\boldsymbol{\varphi}}, \hat{\sigma}_w^2)$
- Yule-Walker estimators:

P_{p+1}^{p}

$$\hat{\boldsymbol{\varphi}} = \hat{\boldsymbol{\Gamma}}_p^{-1} \hat{\boldsymbol{\gamma}}_p, \;\; \hat{\sigma}_w^2 = \hat{\gamma}(0) - \hat{\boldsymbol{\gamma}}_p' \hat{\boldsymbol{\Gamma}}_p^{-1} \hat{\boldsymbol{\gamma}}_p \;\; \Leftrightarrow$$

$$\hat{\boldsymbol{\varphi}} = \hat{\mathbf{R}}_p^{-1} \hat{\boldsymbol{\rho}}_p, \;\; \hat{\sigma}_w^2 = \hat{\gamma}(0) \left[1 - \hat{\boldsymbol{\rho}}_p' \hat{\mathbf{R}}_p^{-1} \hat{\boldsymbol{\rho}}_p \right]$$

where $\hat{\mathbf{R}}_p = \left[\hat{\rho}(k-j) \right]_{k,j=1}^{p}$ & $\hat{\boldsymbol{\rho}}_p' = \left[\hat{\rho}(1) \; \cdots \; \hat{\rho}(p) \right]$

- Equations can be solved recursively using Durbin-Levinson algorithm, giving sample PACF

Large Sample Behavior of Yule-Walker Estimator

- As $n \to \infty$, the Yule-Walker estimators of an AR(p) model behave as:

$$\hat{\boldsymbol{\varphi}} \sim N\left(\boldsymbol{\varphi}, \sigma_w^2 \boldsymbol{\Gamma}_p^{-1} / n\right), \quad \hat{\sigma}_w^2 \to \sigma_w^2$$

  - AR coefficient estimates are asymptotically Normal

- As corollary, sample PACF ($\hat{\varphi}_{hh}$) of AR(p) model for $h>p$ behaves as:

$$\hat{\varphi}_{hh} \sim N\left(\mathbf{0}, 1/n\right), \quad \text{for } h > p$$

  - Result follows since PACF given by Durbin-Levinson solution of Yule-Walker equations

Example

- For AR(2) model with $\hat{\gamma}(0)=1, \hat{\rho}(1)=.5, \& \hat{\rho}(2)=.2$ find Yule-Walker estimates $(\hat{\varphi}, \hat{\sigma}_w^2)$, where $\hat{\varphi} = \begin{bmatrix} \hat{\varphi}_1 \\ \hat{\varphi}_2 \end{bmatrix}$

We have: $\hat{\underline{\Gamma}}_2 \cdot \hat{\varphi} = \hat{\gamma}_2 \iff \hat{\underline{R}}_2 \hat{\varphi} = \hat{\rho}_2$

where $\hat{\underline{R}}_2 = \begin{bmatrix} \hat{\rho}(0) & \hat{\rho}(1) \\ \hat{\rho}(1) & \hat{\rho}(0) \end{bmatrix} = \begin{bmatrix} 1 & .5 \\ .5 & 1 \end{bmatrix}$ & $\hat{\rho}_2 = \begin{bmatrix} \hat{\rho}(1) \\ \hat{\rho}(2) \end{bmatrix} = \begin{bmatrix} .5 \\ .2 \end{bmatrix}$

$\hat{\varphi} = \hat{\underline{R}}_2^{-1} \cdot \hat{\rho}_2 = \begin{bmatrix} 1 & .5 \\ .5 & 1 \end{bmatrix}^{-1} \cdot \begin{bmatrix} .5 \\ .2 \end{bmatrix} = \frac{1}{1-(.5)^2} \cdot \begin{bmatrix} 1 & -.5 \\ -.5 & 1 \end{bmatrix} \cdot \begin{bmatrix} .5 \\ .2 \end{bmatrix} = \ldots = \begin{bmatrix} 16/30 \\ -2/30 \end{bmatrix}$

& $\hat{\sigma}_w^2 = \hat{\gamma}(0) - \hat{\varphi}' \cdot \hat{\gamma}_2 = 1 - \begin{bmatrix} \frac{16}{30} & -\frac{2}{30} \end{bmatrix} \cdot \begin{bmatrix} .5 \\ .2 \end{bmatrix} = 1 - \frac{8-.4}{30} = \frac{22.4}{30}$

Example (cont'd)

- If n=100, find asymptotic distribution of $\hat{\varphi}$

$$\hat{\varphi} \sim N_2\left(\varphi, \hat{\sigma}_w^2 \hat{\Gamma}_2^{-1}/n\right) =$$

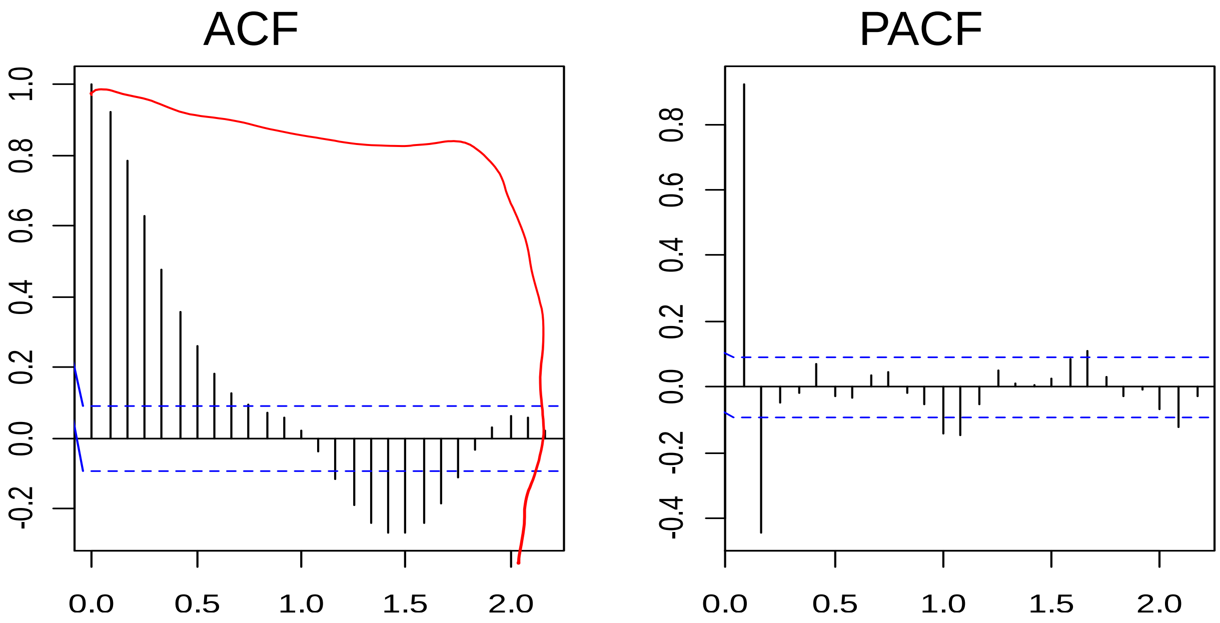
$$= N_2\left(\begin{bmatrix} \varphi_1 \\ \varphi_2 \end{bmatrix}, \frac{22.4}{30} \begin{bmatrix} 4/3 & -2/3 \\ -2/3 & 4/3 \end{bmatrix} \Big/ 100 \right)$$

$\Rightarrow$ Approx 95% Conf. interval for φ_1:

$$\underbrace{\hat{\varphi}_1}: \quad \frac{16}{30} \pm 1.96 \sqrt{\frac{22.4}{30} \cdot \frac{4}{3} \Big/ 100}$$

Example

- Recruitment series:



- → Fit AR(2) model

Yule-Walker Estimation in R

- R function `ar.yw()` performs AR(p) model Yule-Walker estimation
  - E.g. $\underbrace{}$ **`yw.fit = ar.yw( rec, order=2)`**

    data: `rec` — AR order (i.e. p): `order=2`

    - To check estimates:

($\hat{\mu}$) `yw.fit$x.mean` ▶ [1] 62.26278 $\hat{\varphi}_1$

($\hat{\boldsymbol{\varphi}}$) `yw.fit$ar` ▶ [1] 1.3315874 -0.4445447 $\hat{\varphi}_2$

($\hat{\sigma}_w^2$) `yw.fit$var.pred` ▶ [1] 94.79912 σ_w^2

($\hat{\sigma}_w^2 \boldsymbol{\Gamma}_p^{-1} / n$) `yw.fit$asy.var` ▶

```
             [,1]         [,2]
[1,]  0.001783067 -0.001643638
[2,] -0.001643638  0.001783067
```

Forecasting in R

- R function `predict()` builds forecasts
  - E.g. $\texttt{forecast = predict(}\overbrace{\texttt{yw.fit,}}^{\text{fitted model object}}\ \overbrace{\texttt{n.ahead=24}}^{\text{\# of steps ahead}}\texttt{)}$
    - To check forecasts:

			Oct	Nov	...
$(\hat{X}^n_{n+m})$	forecast$pred ►	[1987]	20.62620	26.55461	...
$(\sqrt{P^n_{n+m}})$	forecast$se ►	[1987]	9.736484	16.21387	...

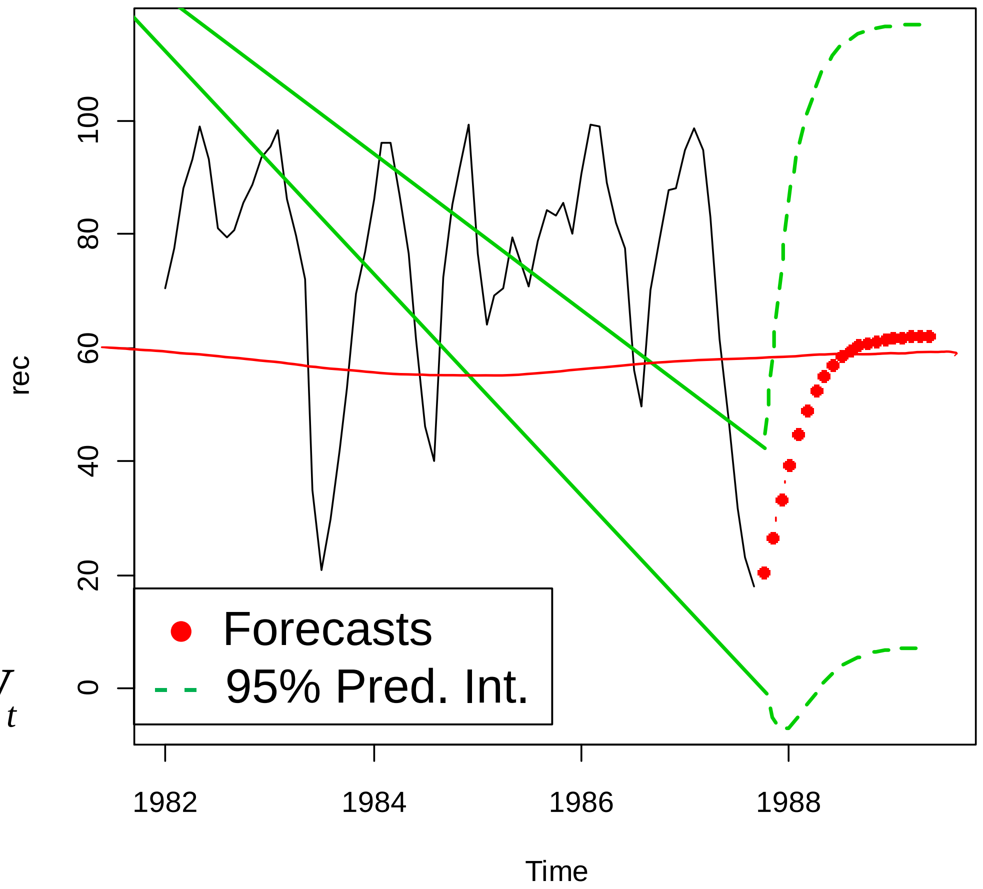
Example

- Recruitment series forecasts based on Yule-Walker AR(2) fitted model:

$$Y_t = \hat{\varphi}_1 Y_{t-1} + \hat{\varphi}_2 Y_{t-2} + W_2$$
$$= 1.3316 Y_{t-1} - .4445 Y_{t-2} + W_t$$

where $Y_t = X_t - \hat{\mu}$ and

$\hat{\mu} = 62.26278, \ \hat{\sigma}_w^2 = 94.8$



Maximum Likelihood AR(p) Estimation

- Instead of Yule-Walker (MM) for AR(p) model, can use Maximum Likelihood (ML) estimation
  - Likelihood is joint density of data $\{x_1,\ldots,x_n\}$ as function of parameters:

$$L\left(\mu,\boldsymbol{\varphi},\sigma_w^2\right)=f\left(x_1,\ldots,x_n;\mu,\boldsymbol{\varphi},\sigma_w^2\right)$$

  - ML estimators are values of parameters that maximize likelihood function
  - What is the likelihood of the AR(p) model?

Maximum Likelihood AR(p) Estimation

- For ML assume AR(p) model is *Gaussian*

$$\left(X_t - \mu\right) = \varphi_1\left(X_{t-1} - \mu\right) + \cdots + \varphi_p (X_{t-p} - \mu) + W_t, \ \underline{\{W_t\} \sim N(0, \sigma_w^2)}$$

- Thus, *conditional* distribution of X_t becomes

$$X_t \mid X_{t-1}, \ldots, X_{t-p} \sim N\left(\mu + \varphi_1\left(X_{t-1} - \mu\right) + \cdots + \varphi_p (X_{t-p} - \mu), \sigma_w^2\right)$$

- This allows us to write likelihood as:

$$L\left(\mu, \boldsymbol{\varphi}, \sigma_w^2\right) = f\left(x_1, \ldots, x_n; \mu, \boldsymbol{\varphi}, \sigma_w^2\right) =$$

$$= f(x_1, x_2, \ldots, x_p) \times f(x_{p+1} \mid x_p, \ldots, x_1) \times \cdots \times f(x_n \mid x_{n-1}, \ldots, x_{n-p})$$

$$= f(x_1, x_2, \ldots, x_p) \times \prod_{t=p+1}^{n} f(x_t \mid x_{t-1}, \ldots, x_{t-p})$$

Maximum Likelihood AR(p) Estimation

- Densities $f(x_t \mid x_{t-1},\ldots,x_{t-p}), \forall t = p+1,\ldots,n$ are trivial to find, but initial density $f(x_1, x_2, \ldots, x_p)$ can be complicated function of parameters.
- Since all densities are Normal, likelihood function will have form:

$$L\left(\mu, \boldsymbol{\varphi}, \sigma_w^2\right) = (2\pi\sigma_w^2)^{-n/2} \times g(\boldsymbol{\varphi}) \times \exp\left[-\frac{S\left(\mu, \boldsymbol{\varphi}\right)}{2\sigma_w^2}\right]$$

  - where S(μ,**φ**) is sum of squares

$$S\left(\mu, \boldsymbol{\varphi}\right) = h(\mu, \boldsymbol{\varphi}) + \sum_{t=p+1}^{n}\left[\left(x_t - \mu\right) - \varphi_1\left(x_{t-1} - \mu\right) - \cdots - \varphi_p\left(x_{t-p} - \mu\right)\right]^2$$

and g(**φ**), h(μ, **φ**) are some functions of μ & **φ**

Example

- Find the likelihood of the AR(1) model:

$$\left(X_t-\mu\right)=\varphi_1\left(X_{t-1}-\mu\right)+W_t,\ \{W_t\}\sim N(0,\sigma_w^2)$$

$$L(\mu,\varphi_1,\sigma_w^2) = f(x_1,\dots,x_n;\mu,\varphi_1,\sigma_w^2) =$$

$$= f(x_1)\times f(x_2|x_1)\times \underbrace{f(x_3|x_2,x_1)}_{f(x_3|x_2)}\times\dots\times \underbrace{f(x_n|x_{n-1},\dots,x_1)}_{f(x_n|x_{n-1})}$$

$$= f(x_1)\times\prod_{t=2}^{n} f(x_t|x_{t-1}),\ \text{where } f(x_t|x_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_w^2}}\cdot e^{-\frac{[(x_t-\mu)-(\varphi x_{t-1}-\mu)]^2}{2\sigma_w^2}}$$

What about $f(x_1)$? To find distr. of X_1, use causal repr.

$$\Rightarrow (X_t-\mu) = \varphi(X_{t-1}-\mu)+W_t = \varphi^2\cdot(X_{t-2}-\mu)+\varphi W_{t-1}+W_t =$$

$$= \dots = \sum_{j=0}^{\infty}\varphi^j\cdot W_{t-j}$$

$$\Rightarrow (X_t - \mu) \sim N\Big(0, \underbrace{Var\Big[\sum_{j=0}^{\infty} \varphi^j W_{t-j}\Big]}\Big)$$

$$= \sum_{j=0}^{\infty} (\varphi^j)^2 \cdot \sigma_w^2 = \frac{\sigma_w^2}{1-\varphi^2}$$

$$\Rightarrow X_1 \sim N\Big(\mu, \frac{\sigma_w^2}{1-\varphi^2}\Big) \Longrightarrow$$

$$\Rightarrow L(\mu, \varphi_1, \sigma_w^2) = \left(\frac{1}{\sqrt{2\pi \frac{\sigma_w^2}{1-\varphi_1^2}}} e^{-\frac{(x_1-\mu)^2}{2\cdot\frac{\sigma_w^2}{1-\varphi_1^2}}}\right) \times \prod_{t=2}^{n} \frac{1}{\sqrt{2\pi\sigma_w^2}} e^{-\frac{[(x_t-\mu)-\varphi_1(x_{t-1}-\mu)]^2}{2\sigma_w^2}}$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma_w^2}}\right)^n \times \underbrace{\sqrt{1-\varphi_1^2}} \times \exp\left\{-\frac{(x_1-\mu)^2\cdot(1-\varphi_1^2) + \sum_{t=2}^{n}[(x_t-\mu)-\varphi_1(x_{t-1}-\mu)]^2}{2\sigma_w^2}\right\}$$

$$\Rightarrow = \left(\frac{1}{\sqrt{2\pi\sigma_w^2}}\right)^n \times g(\varphi) \times \exp\left\{-\frac{S(\mu,\varphi)}{2\sigma_w^2}\right\}$$

where $S(\mu,\varphi) = \underbrace{(1-\varphi_1^2)\cdot(x_1-\mu)^2}_{h(\varphi,\mu)} + \sum_{t=2}^{n}[(x_t-\mu)-\varphi_1(x_{t-1}-\mu)]^2$

Maximum Likelihood AR(p) Estimation

- To find ML estimators $\hat{\mu}, \hat{\boldsymbol{\varphi}}, \hat{\sigma}_w^2$ we need to maximize likelihood $L\left(\mu, \boldsymbol{\varphi}, \sigma_w^2\right)$
  - Can show $\hat{\sigma}_w^2$ always given by $S\left(\hat{\mu}, \hat{\boldsymbol{\varphi}}\right)/n$
  - But there is no closed-form solution for $\hat{\mu}, \hat{\boldsymbol{\varphi}}$
    - That's because of complicated form of $g(\boldsymbol{\varphi}),\ h(\mu, \boldsymbol{\varphi})$
- In practice, use numerical techniques for finding $\hat{\mu}, \hat{\boldsymbol{\varphi}}$ (maximizing L)
  - Common methods are Newton-Raphson or Fisher scoring algorithms

ML Estimation in R

- R function `ar.mle()` performs AR(p) model ML estimation
  - E.g. $\mathtt{\mathbf{ml.fit = ar.mle(}}\ \overbrace{\mathtt{\mathbf{rec}}}^{\text{data}}\mathtt{\mathbf{,}}\ \overbrace{\mathtt{\mathbf{order=2}}}^{\text{AR order (i.e. p)}}\mathtt{\mathbf{)}}$
    - To check estimates:

$(\hat{\mu})$ `ml.fit$x.mean` ▶ `[1] 62.26153`

$(\hat{\boldsymbol{\varphi}})$ `ml.fit$ar` ▶ `[1]  1.3512809 -0.4612736`

$(\hat{\sigma}_w^2)$ `ml.fit$var.pred` ▶ `[1] 89.33597`

$(Cov(\hat{\boldsymbol{\varphi}}))$ `ml.fit$asy.var` ▶

```
             [,1]         [,2]
[1,]  0.001680311 -0.001548918
[2,] -0.001548918  0.001680311
```

Conditional Least Squares AR(p) Estimation

- AR(p) model likelihood is:

$$L\left(\mu, \boldsymbol{\varphi}, \sigma_w^2\right) = f(x_1, x_2, \ldots, x_p) \times \prod_{t=p+1}^{n} f(x_t \mid x_{t-1}, \ldots, x_{t-p})$$

  - What complicates things is initial term $f(x_1, x_2, \ldots, x_p)$
- For large n, effect of initial density is small relative to product $\prod_{t=p+1}^{n} f(x_t \mid x_{t-1}, \ldots, x_{t-p})$
- In this case, can look at *conditional* likelihood

$$L\left(\mu, \boldsymbol{\varphi}, \sigma_w^2 \mid x_1, \ldots, x_p\right) = \prod_{t=p+1}^{n} f(x_t \mid x_{t-1}, \ldots, x_{t-p})$$

  - Condition on first #p values, to remove $f(x_1, x_2, \ldots, x_p)$

Conditional Least Squares AR(p) Estimation

- Conditional likelihood simplifies to

$$L\left(\mu, \boldsymbol{\varphi}, \sigma_w^2 \mid x_1, \dots, x_p\right) = (2\pi\sigma_w^2)^{-\frac{n-p}{2}} \times \exp\left[-\frac{S_c\left(\mu, \boldsymbol{\varphi}\right)}{2\sigma_w^2}\right]$$

  - where $S_c(\mu,\boldsymbol{\varphi})$ is *conditional* sum of squares

$$S_c\left(\mu, \boldsymbol{\varphi}\right) = \sum_{t=p+1}^{n} \left[\left(x_t - \mu\right) - \varphi_1\left(x_{t-1} - \mu\right) - \cdots - \varphi_p\left(x_{t-p} - \mu\right)\right]^2$$

- Maximize conditional likelihood by minimizing $S_c(\mu,\boldsymbol{\varphi})$ using ordinary least square (OLS) from regression
  - In TS, this is called conditional least squares (LS)

Conditional Least Squares AR(p) Estimation

- Conditional LS estimators for AR(p) model are given by OLS estimation of

$$\left(X_t-\hat{\mu}\right)=\hat{\varphi}_1\left(X_{t-1}-\hat{\mu}\right)+\cdots+\hat{\varphi}_p(X_{t-p}-\hat{\mu})\Leftrightarrow$$

$$X_t=\hat{\beta}_0+\hat{\beta}_1\mathrm{X}_{t-1}+\cdots+\hat{\beta}_pX_{t-p},\ \forall p+1\le t\le n$$

  - where: $\hat{\beta}_j=\hat{\varphi}_j, \forall j=1,\dots,p$

$$\hat{\beta}_0=\hat{\mu}\left(1-\hat{\varphi}_1-\cdots-\hat{\varphi}_p\right)$$

$$\hat{\sigma}_w^2=\frac{S_c\left(\hat{\mu},\hat{\boldsymbol{\varphi}}\right)}{n-p}$$

Conditional LS Estimation in R

- R function `ar.ols()` performs AR(p) model conditional ML
  - E.g. $\texttt{ls.fit = ar.ols(}\overbrace{\texttt{rec}}^{\text{data}}\texttt{, }\overbrace{\texttt{order=2}}^{\text{AR order (i.e. p)}}\texttt{)}$
    - To check estimates:

$(\hat{\mu})$ `ls.fit$x.mean` ▶ `[1] 62.26278`

$(\hat{\varphi})$ `ls.fit$ar` ▶ `[1] 1.3540685 -0.4631784`

$(\hat{\sigma}_w^2)$ `ls.fit$var.pred` ▶ `[1] 89.71705`

AR(p) Estimation

- As $n \to \infty$, all AR(p) estimation methods (Yule-Walker, ML, conditional LS) give same results
  - For large n, prefer Yule-Walker / conditional LS
  - They are faster than ML (no need for numerical optimization) and their results are not very different
- However for small sample sizes, and especially if TS is Gaussian, ML estimation performs better than the other two methods
  - For small n, prefer ML estimation

ARMA Estimation

- Have seen estimation for pure AR(p) using:
  - Yule-Walker (MM)
  - Conditional Least Squares (LS)
  - Maximum Likelihood (ML)

better for large n

better for small n

- For estimation in general ARMA model, only rely on ML & LS
  - MM can be problematic / suboptimal for models that contain MA component

Example

- Consider MA(1) model $X_t = W_t + \theta W_{t-1}$
  - MM: estimate (θ, σ_w^2) by matching first two sample & theoretical 2nd order moments

For MA(1): $\begin{cases} \gamma(0) = (1+\theta^2)\sigma_w^2 \\ \gamma(1) = \theta\,\sigma_w^2 \end{cases} \Rightarrow (MM) \begin{cases} \hat\gamma(0) = \hat\sigma_w^2(1+\hat\theta^2) \\ \hat\gamma(1) = \hat\sigma_w^2\,\hat\theta \end{cases} \Rightarrow$

$\Rightarrow \dfrac{\hat\gamma(1)}{\hat\gamma(0)} = \hat\rho(1) = \dfrac{\hat\theta}{1+\hat\theta^2} \Rightarrow \hat\rho(1)\cdot(1+\hat\theta^2) = \hat\theta \Rightarrow$

$\Rightarrow \underbrace{\hat\theta^2 \cdot \hat\rho(1)}_{a} - \underbrace{1}_{b} \cdot \hat\theta + \underbrace{\hat\rho(1)}_{c} = 0 \Rightarrow$ quadratic in $\hat\theta$, has real sol.

iff: $b^2 - 4ac \geq 0 \Rightarrow (-1)^2 - 4\cdot\hat\rho(1)\hat\rho(1) \geq 0$

$\Rightarrow 1 - 4(\hat\rho(1))^2 \geq 0 \Rightarrow (\hat\rho(1))^2 \leq 1/4 \Rightarrow$

$\Rightarrow |\hat\rho(1)| \leq \sqrt{1/4} = 1/2$

In theory $\Rightarrow |\rho(1)| = \left|\dfrac{\theta}{1+\theta^2}\right| \leq \dfrac{1}{2}$

but in practice you could get $|\hat\rho(1)| > \frac{1}{2}$ for some data sets

Maximum Likelihood ARMA Estimation

- Consider Gaussian ARMA(p,q) model:

$$X_t - \mu = \varphi_1\left(X_{t-1} - \mu\right) + \cdots + \varphi_p\left(X_{t-p} - \mu\right) + W_t + \theta_1 W_{t-1} + \cdots + \theta_q W_{t-q}$$

  - Likelihood function: $L\left(\boldsymbol{\beta}, \sigma_w^2\right) = f\left(x_1, \ldots, x_n; \boldsymbol{\beta}, \sigma_w^2\right)$

  with parameter vector $\boldsymbol{\beta} = (\mu, \varphi_1, \ldots, \varphi_p, \theta_1, \ldots, \theta_q)$
- There are two complications with $L(\boldsymbol{\beta})$:
  - WN sequence $\{W_t\}$ is *not observed* $\rightarrow$ cannot use conditional distribution $X_t \mid X_{t-1}, \ldots, X_{t-p}, W_{t-1}, \ldots, W_{t-q} \sim$

$$\sim N\left(\varphi_1(X_{t-1} - \mu) + \cdots + \varphi_p(X_{t-p} - \mu) + W_t + \theta_1 W_{t-1} + \cdots + \theta_q W_{t-q}, \sigma_w^2\right)$$

  - W_t depends on entire past $\{X_{t-1}, \ldots, X_1\}$ (by invertibility)

Maximum Likelihood ARMA Estimation

- Write ARMA likelihood as: $L\left(\boldsymbol{\beta}, \sigma_w^2\right) =$

$$= f(x_1) \times f(x_2 \mid x_1) \times f(x_3 \mid x_2, x_1) \times \cdots \times f\left(x_n \mid x_{n-1}, \ldots, x_1\right)$$

$$\Rightarrow L\left(\boldsymbol{\beta}, \sigma_w^2\right) = \prod_{t=1}^{n} f\left(x_t \mid x_{t-1}, \ldots, x_1\right)$$

  - For Gaussian ARMA models, $X_t \mid X_{t-1}, \ldots, X_1$ is equal to BLP X_t^{t-1}, which is Normally distributed (as linear function of Normals) with variance P_t^{t-1} !!

$$X_t \mid X_{t-1}, \ldots, X_1 \sim N\left(X_t^{t-1}, P_t^{t-1}\right) \Rightarrow$$

$$\Rightarrow f\left(x_t \mid x_{t-1}, \ldots, x_1\right) = \frac{1}{\sqrt{2\pi P_t^{t-1}}} \exp\left[-\frac{(x_t - x_{t-1}^{t})^2}{2P_t^{t-1}}\right]$$

Maximum Likelihood ARMA Estimation

- Given data $\{x_1, \ldots, x_n\}$, every BLP x_{t-1}^t is function of the parameters $\boldsymbol{\beta} \Rightarrow$ write $x_{t-1}^t(\boldsymbol{\beta})$
- Also $P_t^{t-1} = \gamma(0)\prod_{j=1}^{t-1}(1-\varphi_{jj}^2) = \sigma_w^2\, r_t(\boldsymbol{\beta})$, where $r_{t+1}(\boldsymbol{\beta}) = (1-\varphi_{tt}^2) r_t(\boldsymbol{\beta})$ & $r_1(\boldsymbol{\beta}) = \gamma(0)/\sigma_w^2 = \sum_{j=0}^{\infty}\psi_j^2$
- Thus, $L(\boldsymbol{\beta}, \sigma_w^2) = \prod_{t=1}^{n} f\left(x_t \mid x_{t-1}, \ldots, x_1\right) =$

$$= (2\pi\sigma_w^2)^{-n/2}\left[r_1(\boldsymbol{\beta}) \times r_2(\boldsymbol{\beta}) \times \cdots \times r_n(\boldsymbol{\beta})\right]^{-1/2} \exp\left[-\frac{S(\boldsymbol{\beta})}{2\sigma_w^2}\right]$$

  - where: $S(\boldsymbol{\beta}) = \sum_{t=1}^{n} \frac{[x_t - x_{t-1}^t(\boldsymbol{\beta})]^2}{r_t(\boldsymbol{\beta})}$

Maximum Likelihood ARMA Estimation

- To find ML estimators $\hat{\boldsymbol{\beta}}, \hat{\sigma}_w^2$ we need to maximize likelihood $L(\boldsymbol{\beta}, \sigma_w^2)$
  - Can show $\hat{\sigma}_w^2$ always given by $S(\hat{\boldsymbol{\beta}}) / n$
  - There is no closed-form solution for $\hat{\boldsymbol{\beta}}$
- In practice, use numerical techniques for finding $\hat{\boldsymbol{\beta}}$ (maximizing L)
  - Common methods are Newton-Raphson or Fisher scoring algorithms

Conditional Least Squares ARMA Estimation

- Likelihood function can be complicated
  - Need 1-step-ahead BLP $X_t^{t-1}, \forall t = 1, \ldots, n$
- Can simplify estimation by conditioning on first p values & using truncated prediction:

$$w_t(\boldsymbol{\beta}) = (x_t - \mu) - \sum_{j=1}^{p} \varphi_j (x_{t-j} - \mu) - \sum_{k=1}^{q} \theta_k w(\boldsymbol{\beta})_{t-k}, \ \ \forall t \geq p+1$$

  where $w_p(\boldsymbol{\beta}) = w_{p-1}(\boldsymbol{\beta}) = \ldots = w_{1-q}(\boldsymbol{\beta}) = 0$
- Estimate $\hat{\boldsymbol{\beta}}$ by minimizing conditional sum of squares $S_c(\boldsymbol{\beta}) = \sum_{t=p+1}^{n} w_t^2(\boldsymbol{\beta})$ and $\hat{\sigma}_w^2$ by $S_c(\hat{\boldsymbol{\beta}}) / n - p$

Large Sample Behavior of ARMA Estimation

- As n→∞, ARMA estimators (either ML or LS) behave as: $(\hat{\boldsymbol{\varphi}},\hat{\boldsymbol{\theta}})' \sim N\left((\boldsymbol{\varphi},\boldsymbol{\theta})', \sigma_w^2 \boldsymbol{\Gamma}_{p,q}^{-1} \big/ n\right),\ \hat{\sigma}_w^2 \to \sigma_w^2$

  - where $\boldsymbol{\Gamma}_{p,q} = \begin{bmatrix} \boldsymbol{\Gamma}_{\boldsymbol{\varphi\varphi}} & \boldsymbol{\Gamma}_{\boldsymbol{\varphi\theta}} \\ \boldsymbol{\Gamma}_{\boldsymbol{\theta\varphi}} & \boldsymbol{\Gamma}_{\boldsymbol{\theta\theta}} \end{bmatrix}$ and

$\boldsymbol{\Gamma}_{\boldsymbol{\varphi\varphi}} = \left[\gamma_Y(i-j)\right]_{i,j=1}^{p}$ for AR(p) model: $\varphi(B)Y_t = W_t$

$\boldsymbol{\Gamma}_{\boldsymbol{\theta\theta}} = \left[\gamma_Z(i-j)\right]_{i,j=1}^{q}$ for AR(q) model: $\theta(B)Z_t = W_t$

$\boldsymbol{\Gamma}_{\boldsymbol{\varphi\theta}} = \boldsymbol{\Gamma}'_{\boldsymbol{\theta\varphi}} = \left[\gamma_{YZ}(i-j)\right]_{i=1,j=1}^{p,q}$ for cross-covariance $\gamma_{YZ}(h)$

(Note: $\boldsymbol{\Gamma}_{\boldsymbol{\varphi\varphi}}$ is $p \times p$, $\boldsymbol{\Gamma}_{\boldsymbol{\theta\theta}}$ is $q \times q$, and $\boldsymbol{\Gamma}_{\boldsymbol{\varphi\theta}}$ is $p \times q$)

Example

$$\Rightarrow \Gamma_{p,q} = \sigma_w^2 \begin{bmatrix} 1/1-\varphi^2 & 1/1+\varphi\theta \\ 1/1+\varphi\theta & 1/1-\theta^2 \end{bmatrix}$$

$$X_t - \varphi X_{t-1} = W_t + \theta W_{t-1}$$

- Find $\sigma_w^2 \Gamma_{p,q}^{-1}$ for ARMA(1,1) model

$$\Gamma_{p,q} = \begin{bmatrix} \Gamma_{\varphi\varphi} & \Gamma_{\varphi\theta} \\ \Gamma_{\theta\varphi} & \Gamma_{\theta\theta} \end{bmatrix}, \text{ where } \begin{cases} \Gamma_{\varphi\varphi} = [\gamma_Y(i-j)]_{i,j=1}^{1} = \gamma_Y(0) = Var[Y_t] \\ Y_t - \varphi Y_{t-1} = W_t \Rightarrow Var[Y_t] = \boxed{\sigma_w^2 \frac{1}{1-\varphi^2} = \Gamma_{\varphi\varphi}} \end{cases}$$

$$\Gamma_{\theta\theta} = [\gamma_Z(i-j)]_{i,j=1}^{1} = \gamma_Z(0) = Var[Z_t], \text{ where } Z_t = -\theta Z_{t-1} + W_t$$

$$\Rightarrow \gamma_Z(0) = \dots = \boxed{\sigma_w^2 \frac{1}{1-\theta^2} = \Gamma_{\theta\theta}}$$

$$\Gamma_{\varphi\theta} = [\gamma_{YZ}(i,j)]_{i,j=1}^{1} = \gamma_{YZ}(1,1) = Cov(Y_t, Z_t) = Cov(\varphi Y_{t-1} + W_t, -\theta Z_{t-1} + W_t) =$$

$$= -\varphi\theta \underbrace{Cov(Y_{t-1}, Z_{t-1})}_{\gamma_{YZ}(1,1)} - \theta \underbrace{Cov(W_t, Z_{t-1})}_{0} + \varphi \underbrace{Cov(Y_{t-1}, W_t)}_{0} + \underbrace{Cov(W_t, W_t)}_{=\sigma_w^2}$$

$$\Rightarrow \Gamma_{\varphi\theta} = \sigma_w^2 / 1+\varphi\theta$$

ML Estimation in R

- R function `arima()` performs ARMA estimation

For ML estimation:

$$\texttt{ml.fit = arima(} \underbrace{\texttt{soi,}}_{\text{data}} \underbrace{\texttt{order=c(2,0,2)}}_{\text{ARMA order as (p,0,q)}} \texttt{)}$$

- To check estimates:

$(\hat{\boldsymbol{\beta}})$ **ml.fit$coef** ▶

```
       ar1         ar2         ma1         ma2   intercept
1.66444102 -0.92137814 -1.40571556  0.79169887  0.08177904
```

$(\hat{\sigma}_w^2)$ **ml.fit$sigma2** ▶ `[1] 0.08657299`

$(Cov(\hat{\boldsymbol{\beta}}))$ **ml.fit$var.coef** ▶

```
                    ar1           ar2           ma1           ma2     intercept
ar1        7.760201e-04 -7.019119e-04 -9.597996e-04  4.587052e-04 -2.718178e-06
ar2       -7.019119e-04  7.568129e-04  1.038739e-03 -6.326237e-04  3.192218e-06
ma1       -9.597996e-04  1.038739e-03  2.539872e-03 -1.711167e-03  9.465564e-06
ma2        4.587052e-04 -6.326237e-04 -1.711167e-03  1.539701e-03 -3.490782e-06
intercept -2.718178e-06  3.192218e-06  9.465564e-06 -3.490782e-06  4.318156e-04
```

LS Estimation in R

- For large n, perform conditional LS estimation
  - Similar results to ML, but simpler (→faster) to run

conditional LS: minimize Conditional Sum of Squares (CSS)

```
ls.fit = arima( soi, order=c(2,0,2), method="CSS" )
```

- To check estimates:

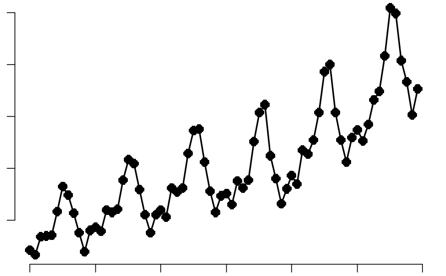
$(\hat{\boldsymbol{\beta}})$ `ls.fit$coef` ►

```
       ar1         ar2         ma1         ma2   intercept 
1.66075580 -0.91670817 -1.38960886  0.78045434  0.08296247
```

$(\hat{\sigma}_w^2)$ `ls.fit$sigma2` ► `[1] 0.08721947`

$(Cov(\hat{\boldsymbol{\beta}}))$ `ls.fit$var.coef` ►

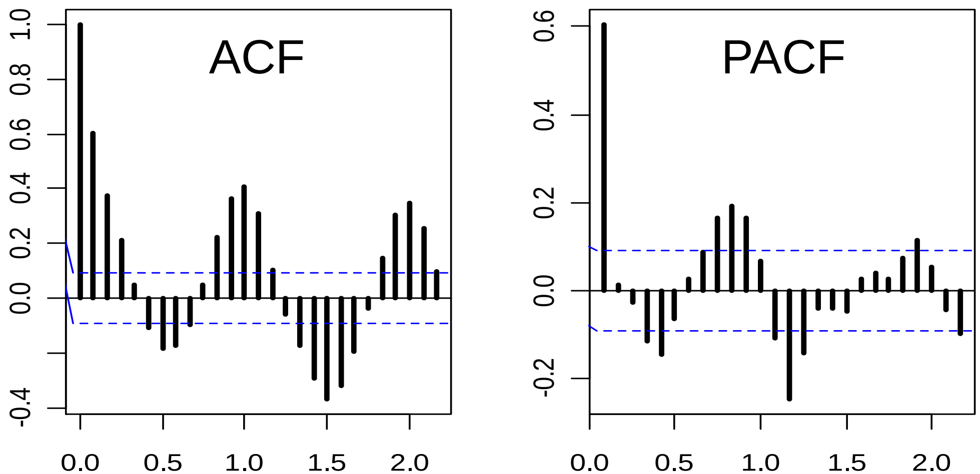
```
                    ar1           ar2           ma1           ma2     intercept
ar1        8.343171e-04 -7.425858e-04 -8.941348e-04  0.0003802571  4.195686e-06
ar2       -7.425858e-04  7.941902e-04  9.599489e-04 -0.0005672521 -5.946319e-06
ma1       -8.941348e-04  9.599489e-04  2.227674e-03 -0.0015083161 -1.183077e-05
ma2        3.802571e-04 -5.672521e-04 -1.508316e-03  0.0014779296  1.684810e-05
intercept  4.195686e-06 -5.946319e-06 -1.183077e-05  0.0000168481  4.468326e-04
```



6. ARMA Model Selection

Model Selection

- Often, the ARMA order (i.e. the p,q values), is not obvious from ACF & PACF plots
  - E.g.



- Using high (p,q) order leads to *suboptimal estimation* (overfitting)

Example

For AR(p): $\hat{\varphi} \sim N(\varphi, \sigma_w^2 \cdot \Gamma_p^{-1}/n)$ (as $\underline{n} \to \infty$)

Also, for $\underline{AR(1)}$: $\gamma(0) = \frac{\sigma_w^2}{1-\varphi^2}$ & $\gamma(h) = \varphi^h \cdot \gamma(0)$, $\forall h \geq 1$

- Assume AR(1) data, and compare $\hat{\varphi}_1$ asymptotic variance from fitting AR(1) & AR(2) models

$\underline{\text{For AR(1) fit:}}$ $\hat{\varphi}_1 \sim N(\varphi_1, \frac{\sigma_w^2}{n} \cdot \Gamma_1^{-1})$ where $\Gamma_1 = \gamma(0) = \frac{\sigma_w^2}{1-\varphi^2} \Rightarrow$

$\Rightarrow \hat{\varphi}_1 \sim N(\varphi_1, \frac{1-\varphi^2}{n}) \Rightarrow s.e.(\hat{\varphi}_1) = \sqrt{\frac{1-\varphi^2}{n}}$

$\underline{\text{For AR(2) fit:}}$ $\hat{\varphi}_2 = \begin{bmatrix} \hat{\varphi}_1 \\ \hat{\varphi}_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \varphi_1 \\ 0 \end{bmatrix}, \frac{\sigma_w^2}{n} \cdot \Gamma_2^{-1}\right)$, where

$\Gamma_2 = \begin{bmatrix} \gamma(0) & \gamma(1) \\ \gamma(1) & \gamma(0) \end{bmatrix} = \gamma(0) \begin{bmatrix} 1 & \varphi_1 \\ \varphi_1 & 1 \end{bmatrix} = \frac{\sigma_w^2}{1-\varphi_1^2} \begin{bmatrix} 1 & \varphi_1 \\ \varphi_1 & 1 \end{bmatrix} \Rightarrow \Gamma_2^{-1} = \frac{1-\varphi_1^2}{\sigma_w^2} \cdot \begin{bmatrix} 1 & \varphi_1 \\ \varphi_1 & 1 \end{bmatrix}^{-1} =$

$= \frac{\cancel{1-\varphi_1^2}}{\sigma_w^2} \cdot \frac{1}{\cancel{1-\varphi_1^2}} \begin{bmatrix} 1 & -\varphi_1 \\ -\varphi_1 & 1 \end{bmatrix}$

$\Rightarrow \hat{\varphi}_2 \sim N\left(\begin{bmatrix} \varphi_1 \\ 0 \end{bmatrix}, \frac{\cancel{\sigma_w^2}}{n} \cdot \frac{1}{\cancel{\sigma_w^2}} \begin{bmatrix} 1 & -\varphi_1 \\ -\varphi_1 & 1 \end{bmatrix}\right) \Rightarrow s.e.(\hat{\varphi}_1) = \frac{1}{\sqrt{n}}$

Model Selection

- Want to find simplest possible ARMA(p,q) model to describe data
- To compare models, optimize some criterion that balances fit & complexity
  - *Akaike's Information Criterion* (AIC):

$$AIC(p,q) = -2\log L(\hat{\boldsymbol{\beta}}, \hat{\sigma}_w^2) + 2(p+q+1)$$

  - *Bayesian Information Criterion* (BIC):

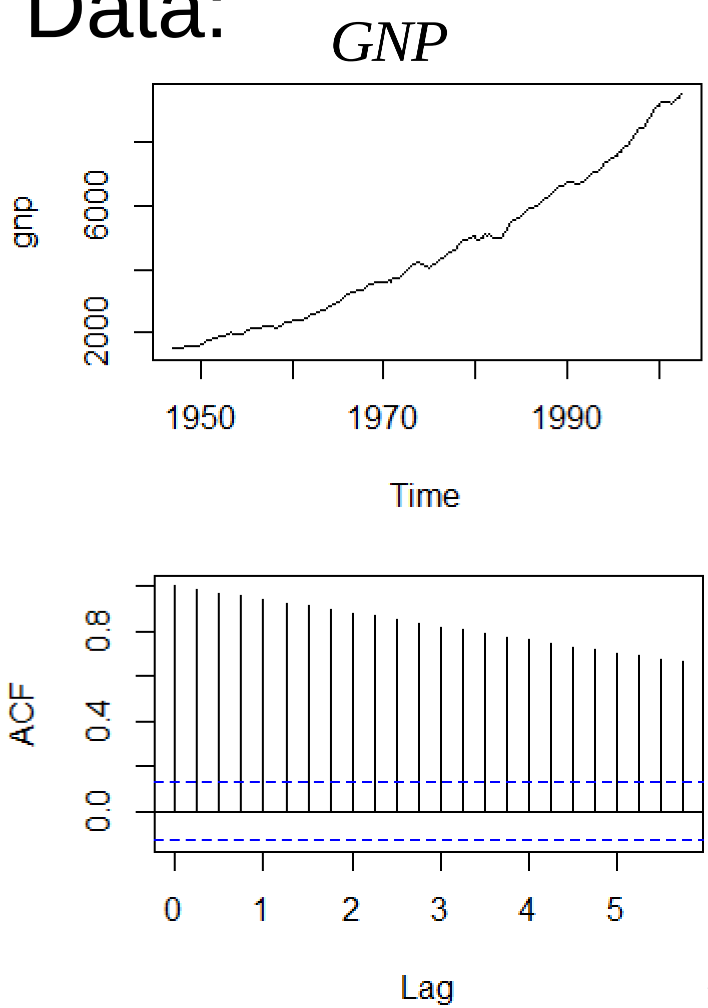
$$BIC(p,q) = \log \hat{\sigma}_w^2 + \frac{(p+q+1)\log n}{n}$$

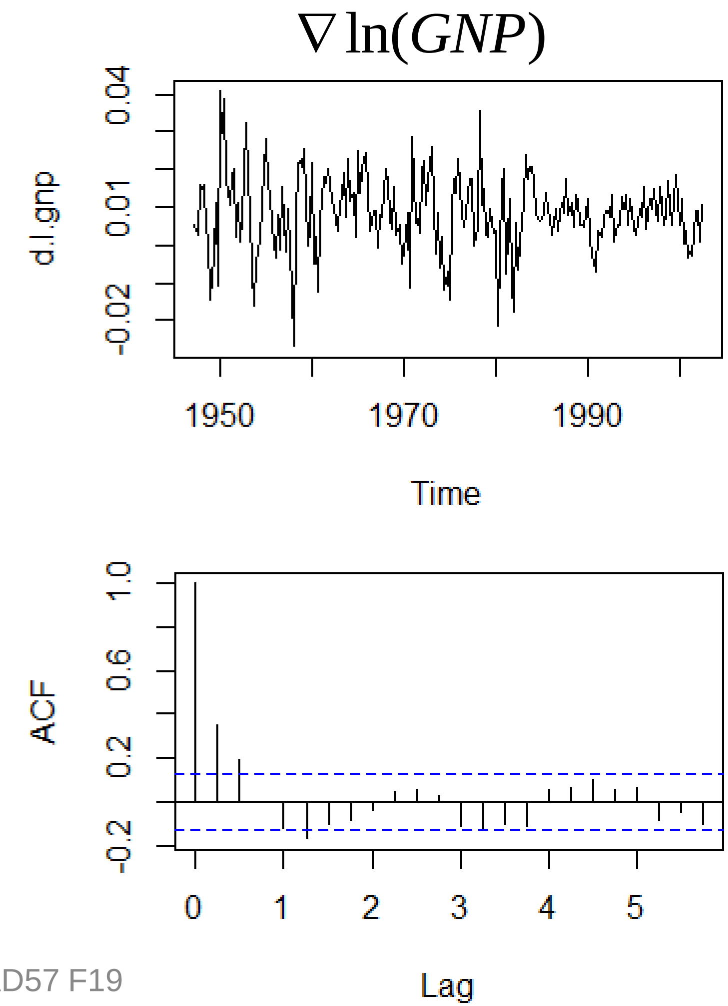
Model Selection

- Fit models with different (p,q) orders, and select one that has the *minimum* AIC / BIC
- In general, BIC tends to give simpler (i.e. smaller) models than AIC
  - BIC puts higher penalty on # of model parameters
- In practice, use AIC if model is meant for making predictions, and BIC if model is meant for describing TS mechanics

Example

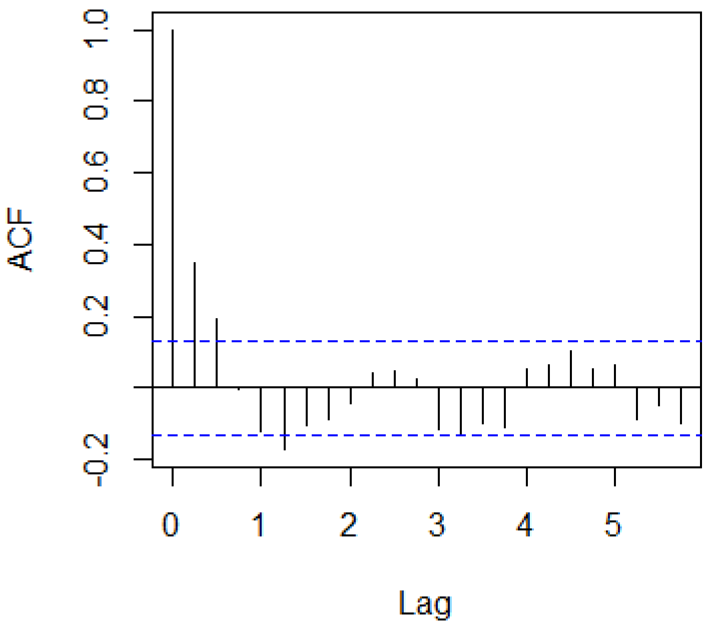
- Data:

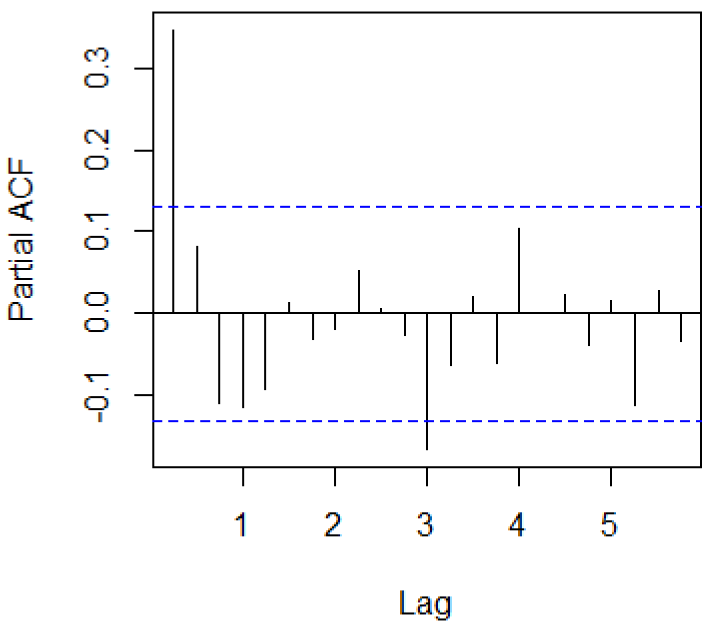




Example (cont'd)

- ARMA order for ∇ln(GNP) ?





Example (cont'd)

- Compare ARMA(0,2) & ARMA(1,0) models for $\nabla \ln(\text{GNP})$ data

```
fit_ar = arima(d.l.gnp, order=c(1,0,0))
fit_ma = arima(d.l.gnp ,order=c(0,0,2))
```

```
AIC(fit_ar) ► -1431.221
AIC(fit_ma) ► -1431.929
```

based on AIC, select ARMA(0,2)

```
BIC(fit_ar) ► -1421.013
BIC(fit_ma) ► -1418.319
```

based on BIC, select ARMA(1,0)

Model Selection

- Following `R` functions automatically compare competing models:
  - Select best AR(p) from p=1,…,order.max:

    `ar(d.l.gnp, order.max=20, aic=TRUE)`

    $\rightarrow$ gives AR(4) w/ AIC = −1432.47
  - Forward selection of best ARMA(p,q) model

    `auto.arima(d.l.gnp)`

    $\rightarrow$ gives AR(1) w/ AIC = −1431.22

Diagnostics

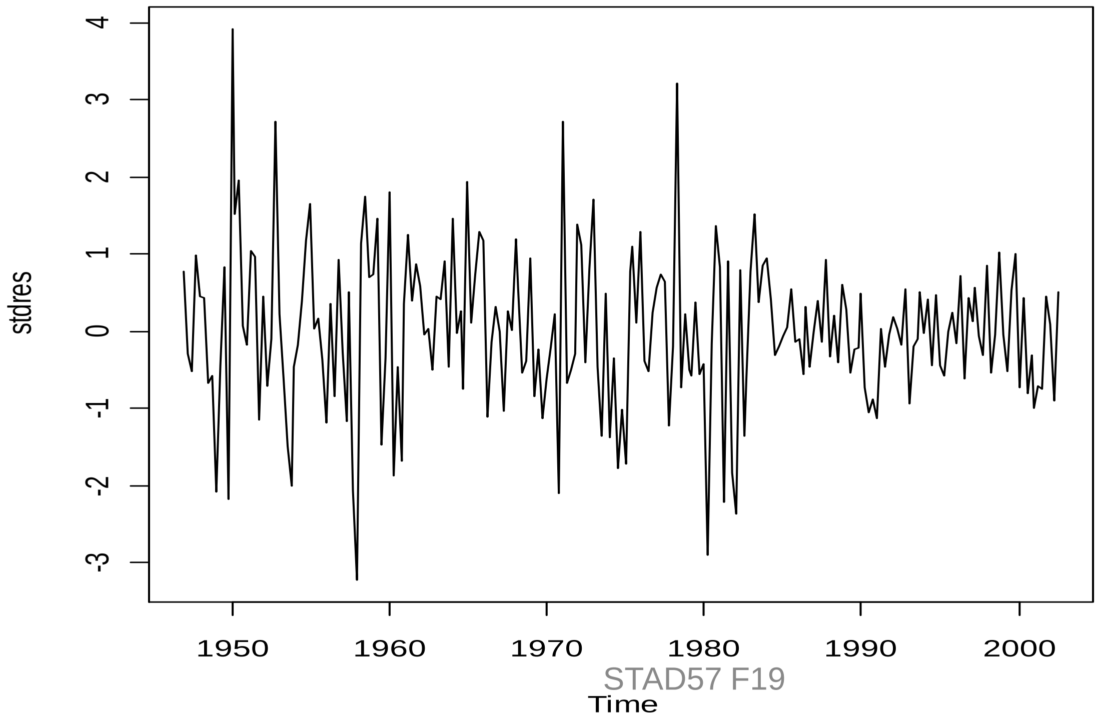
- After fitting an ARIMA model to data, perform diagnostics checks to asses the fit
- As usual, diagnostics rely on model *residuals* For TS models, residuals are essentially the (estimated) innovations $r_t = X_t - \hat{X}_t^{t-1}$
  - *Standardized residuals* given by:

$$e_t = \left(X_t - \hat{X}_t^{t-1} \right) \Big/ \sqrt{\hat{P}_t^{t-1}}$$

  - $\{e_t\}$ should follow standard Normal White Noise

Diagnostics

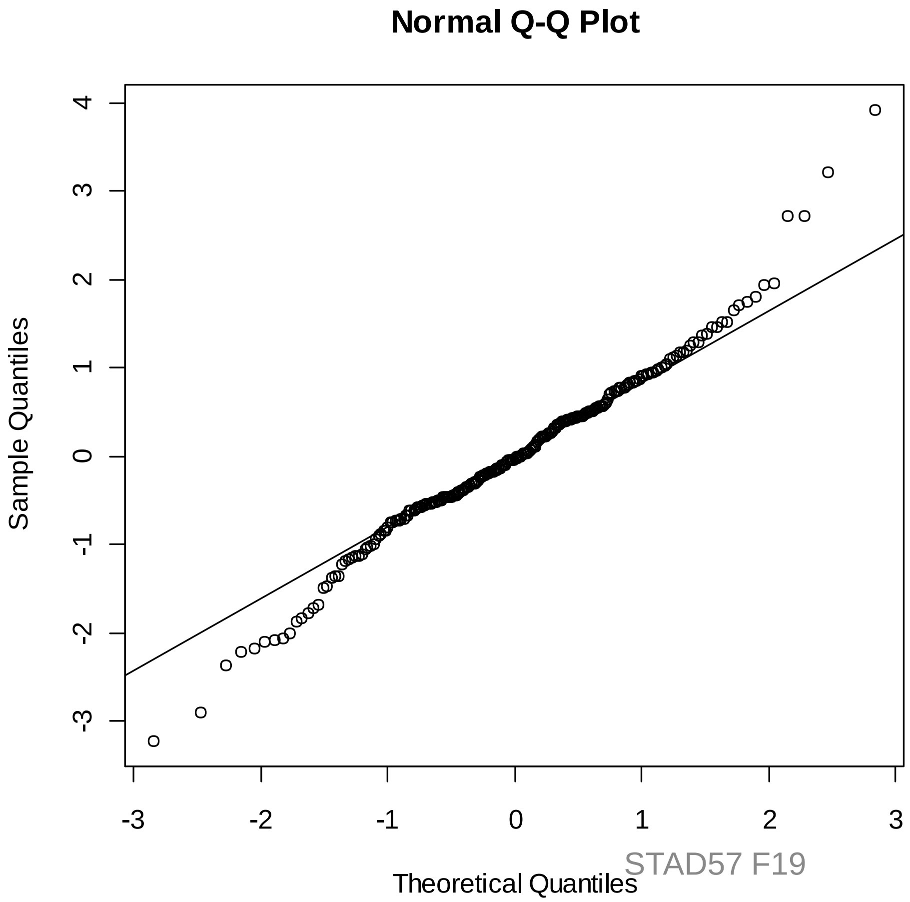
- To check against changes in mean / variance level, plot $\{e_t\}$
  - E.g. ARMA(0,2) for ∇ln(GNP) series



If $\{e_t\}$ don't have 0-mean & constant variance, perhaps some preprocessing (de-trending and/or transformation) is needed

Diagnostics

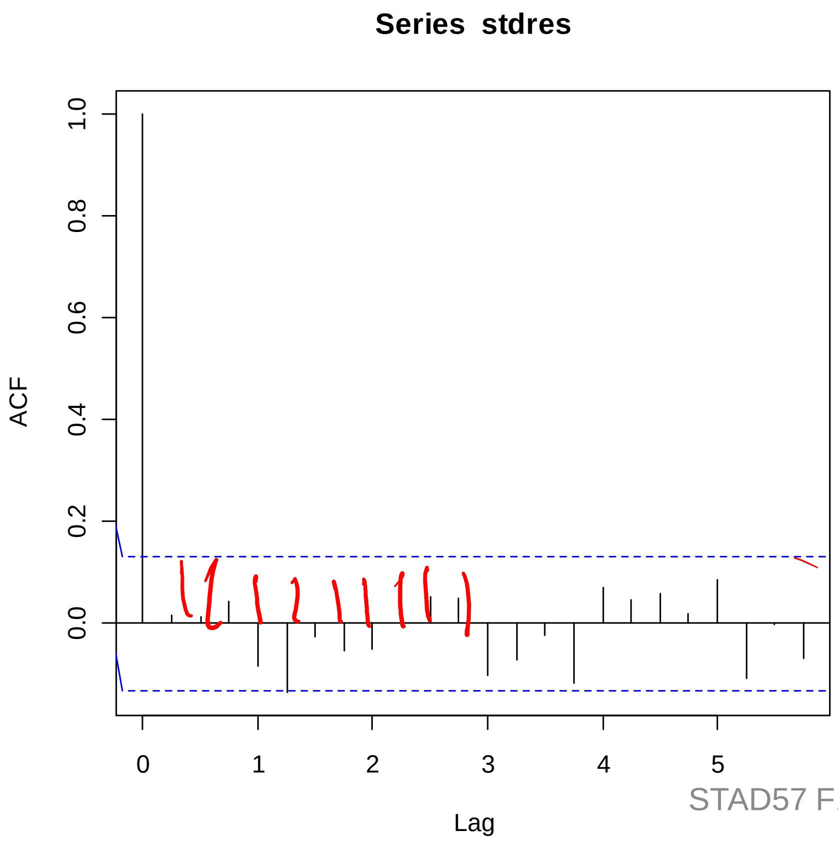
- To check against Normality, look at QQ-plot



Deviations from Normality degrade the quality of the estimates & their approximate distribution, especially for small sample sizes. In some cases, transformations can help

Diagnostics

- To check against remaining auto-dependence not captured by model, look at *residual* ACF



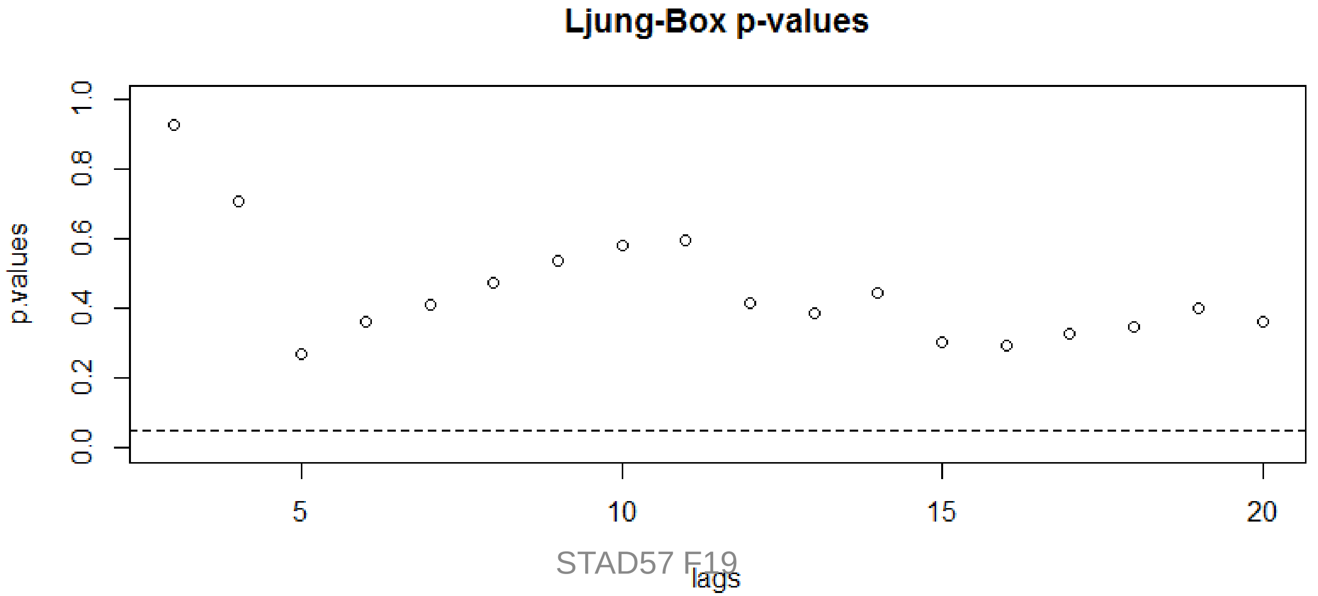
If there are significant auto-correlations in the residuals, the model could be refined, by increasing AR and/or MA order

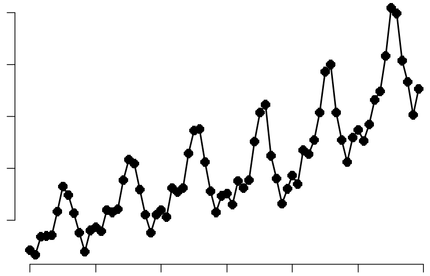
Diagnostics

- Common method for *testing* against residual auto-correlation is Ljung–Box test:
  - For WN series, sample ACF $\hat{\rho}_e(h) \sim N\left(0,1/n\right), \ \forall h$
    - Also know that sum of squared standard Normals follows χ^2 (chi-square) distribution
  - Ljung-Box statistic: $Q = n(n+2)\sum_{h=1}^{H} \frac{\hat{\rho}_e^2(h)}{n-h}$
  - If $\{e_t\}$~WN, then $Q \sim \chi^2_{H-p-q}$, as $n \to \infty$
  - Used to *simultaneously* test if first $\#H$ auto-correlations are equal to 0 (usually H=20)

Diagnostics

- Often, series of Ljung-Box tests are performed for a range of H's & their p-values are plotted
  - If residuals are independent, p-values should be high, e.g. >5% (WN null hypothesis not rejected)





7. Integrated & Seasonal ARMA Models

Integrated Series

- Can extend ARMA framework to deal with (non-stationary) *random walk* type models
  - E.g. $X_t = X_{t-1} + W_t,\ \{W_t\} \sim WN(0, \sigma_w^2)$
  - We know that 1st order differences of random walk are *stationary* $\rightarrow$ $\nabla X_t = (1-B) X_t = W_t$
- Series is called *integrated of order d,* if its dth order differences are stationary:

$\{X_t\}$ non-stationary, but $\{\nabla^d X_t = (1-B)^d X_t\}$ stationary

(and $\{\nabla^b X_t\}$ non-stationary $\forall b < d$)

Examples

- Integrated of order 1 (random walk with drift):

$$X_t = \mu + X_{t-1} + W_t, \ \{W_t\} \sim WN(0, \sigma_w^2)$$

$$\nabla X_t = X_t - X_{t-1} = (\mu + \cancel{X_{t-1}} + W_t) - \cancel{X_{t-1}} = \mu + W_t \sim WN(\mu, \sigma_w^2)$$

- Integrated of order 2:

$$X_t = X_{t-1} + Z_t, \ \begin{cases} Z_t = Z_{t-1} + W_t \\ \{W_t\} \sim WN(0, \sigma_w^2) \end{cases}$$

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

$$= (\cancel{X_{t-1}} + Z_t) - \cancel{2}X_{t-1} + X_{t-2} = Z_t - (\cancel{X_{t-2}} + Z_{t-1}) + \cancel{X_{t-2}} =$$

$$= Z_t - Z_{t-1} = W_t \sim WN(0, \sigma_w^2)$$

ARIMA Models

- Series follows *integrated ARMA*(p,q) model of order d, or *ARIMA*(p,d,q), if:

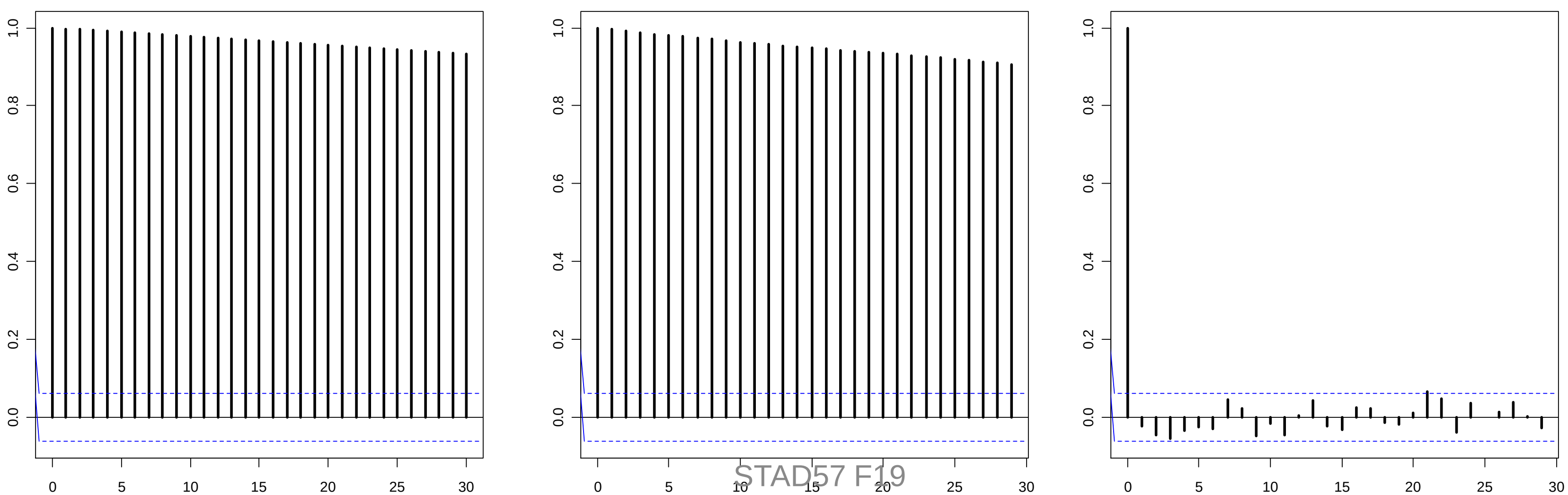
$$\{\nabla^d X_t\} \sim \text{ARMA}(p,q) \Leftrightarrow \varphi(B)(1-B)^d X_t = \theta(B)W_t$$

  - Where $W_t \sim WN(0,\sigma_w^2)$, $\varphi(B) = 1 - \varphi_1 B - \cdots - \varphi_p B^p$, and $\theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$
  - If $E\left[\nabla^d X_t\right] = \mu \neq 0$, write model as:

$$\varphi(B)\left(\nabla^d X_t - \mu\right) = \theta(B)W_t$$

ARIMA Models

- To identify integration order, plot ACF for original & differenced series, and use lowest differencing order that is stationary
  - E.g. for ARIMA of order d=2
    - $\mathrm{ACF}(X_t)$
    - $\mathrm{ACF}(\nabla X_t)$
    - $\mathrm{ACF}(\nabla^2 X_t)$



ARIMA Models

Two things to note about ARIMA models:

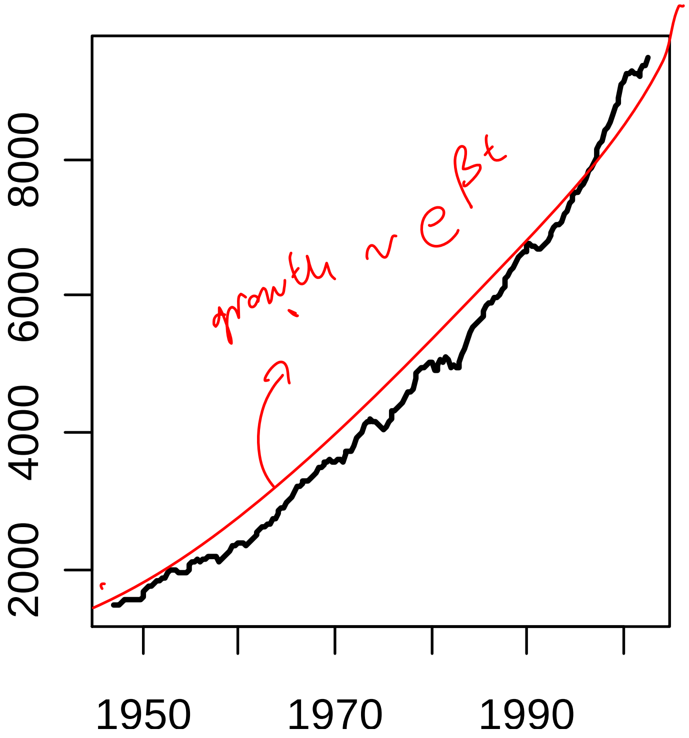
- Over-differencing can introduce *artificial* dependencies; e.g. for $X_t = X_{t-1} + W_t$ we have:
  - $\nabla X_t = W_t \sim WN,$ but $\nabla^2 X_t = W_t - W_{t-1} \sim \mathrm{MA}(1)$
- ARIMA forecasts behave different than ARMA
  - E.g. $X_t \sim \mathrm{ARIMA}(p,1,q) \Rightarrow Y_t = \nabla X_t \sim \mathrm{ARMA}(p,q)$

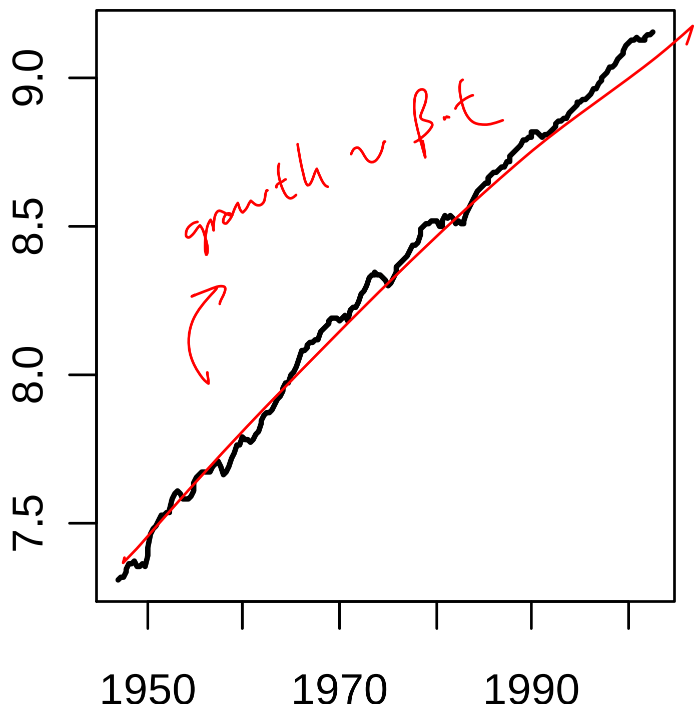
$\left(Y_{n+m}^n, P_{n+m}^n(Y)\right) \to \left(\mu, \gamma(0)\right)$ as $m \to \infty$

but $\begin{cases} X_{n+m}^n = X_{n+m-1}^n + Y_{n+m}^n \to \pm\infty \text{ (if } \mu \neq 0) \\ P_{n+m}^n(X) \to \infty \end{cases}$, as $m \to \infty$

Example

- US quarterly GNP
  - Actual X_t
  - Log-transform $Y_t = \log(X_t)$





(logarithmic transform removes exponential trend)

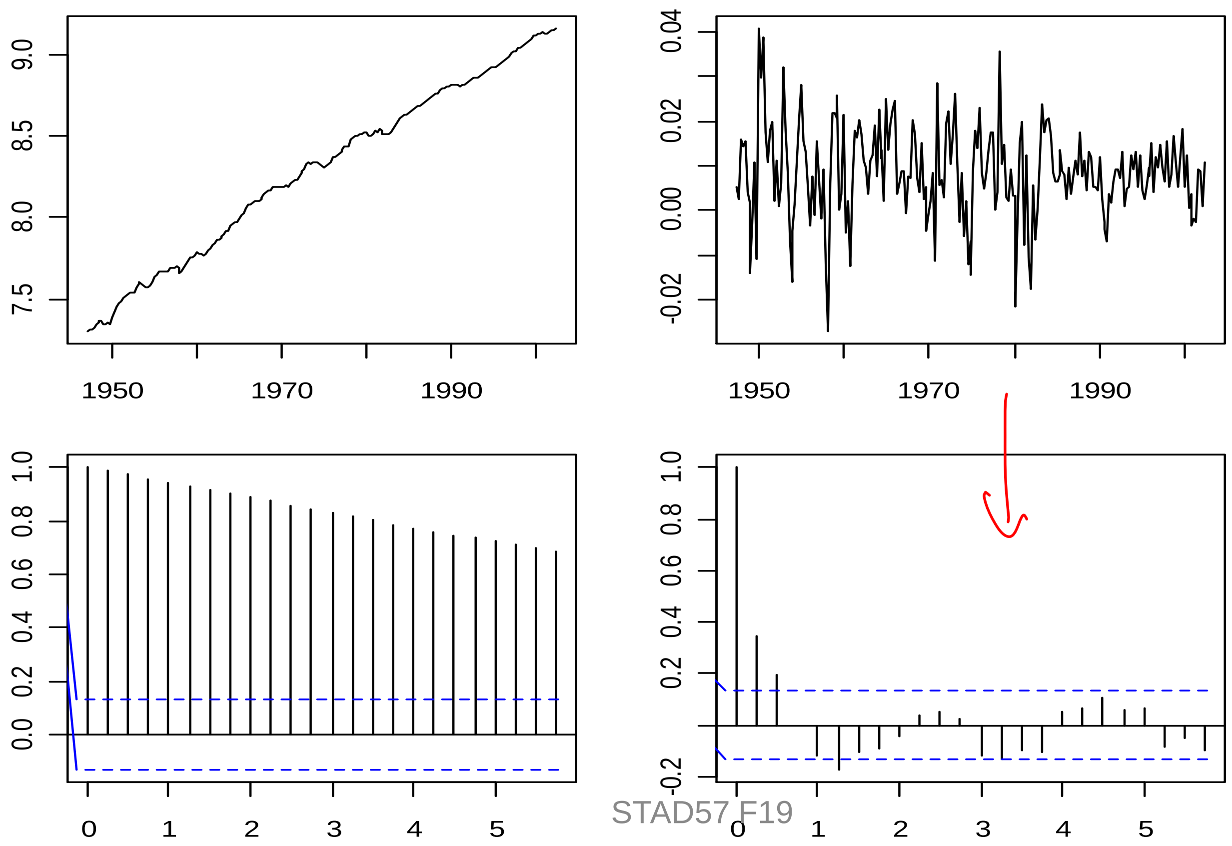
Example

- Find integration order

$\bullet\, Y_t$ $\bullet\, \nabla Y_t$

Series

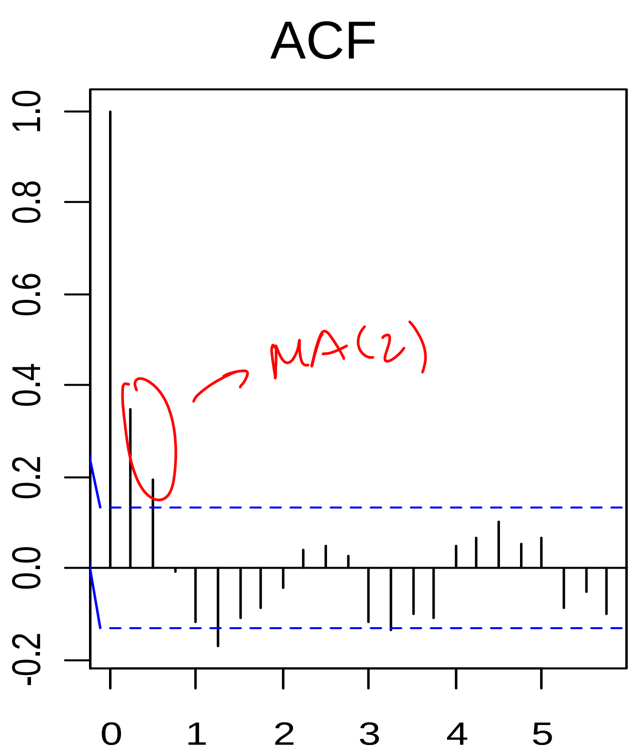
ACF

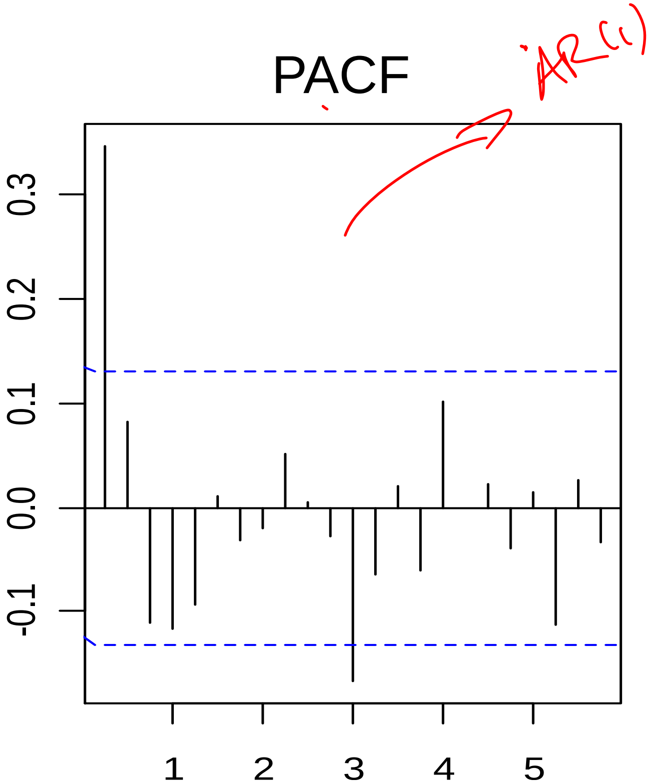


$\Rightarrow d = 1$

Example

- Choose ARMA order (i.e. p,q) for ∇Y_t





Example

p = AR order
d = order of integration
q = MA order

- Fit ARIMA(0,1,2) model in R

$$\texttt{fit\_ma = sarima(}\underbrace{\texttt{ log(gnp)}}_{\text{data}}\texttt{, }\underbrace{\texttt{0, 1, 2}}_{\text{(p,d,q)}}\texttt{)}$$

in package "astsa" (from textbook)

- To check estimates:

```
fit_ma ▶
Coefficients:
           ma1      ma2   constant
        0.3028   0.2036     0.0083
s.e.    0.0654   0.0644     0.0010

sigma^2 estimated as 8.919e-05:
log likelihood = 719.96
```

Example

- Forecasts from ARIMA(0,1,2)

data | steps ahead | (p,d,q)

```
fore = sarima.for(log(gnp), n.ahead=20, 0, 1, 2)
```

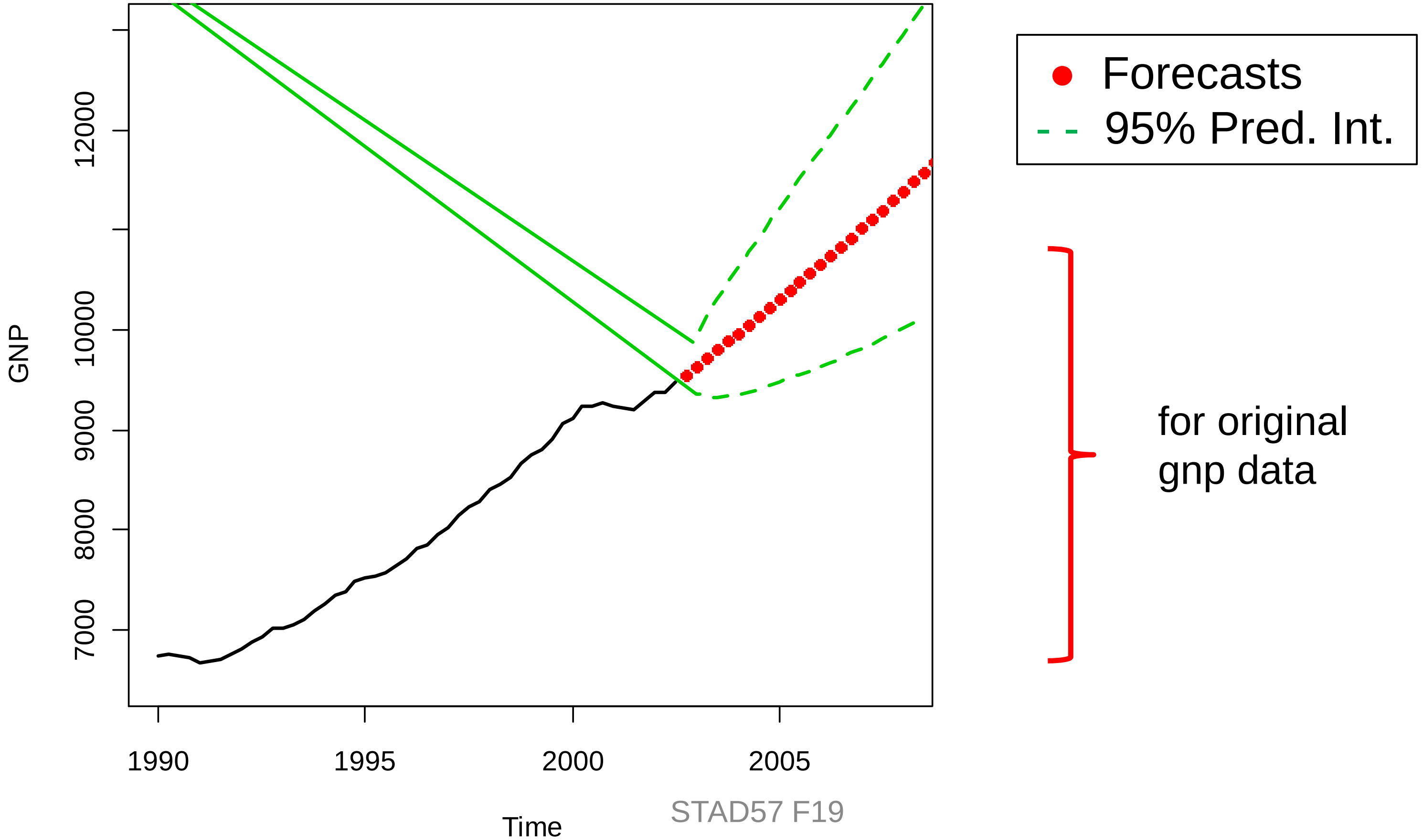
- To check forecasts:

```
fore$pred ►
          Qtr1     Qtr2     Qtr3     Qtr4
2002                            9.164779
2003 9.174085 9.182415 9.190745 9.199075
2004 9.207405 9.215735 9.224064 9.232394

fore$se ►
            Qtr1        Qtr2        Qtr3        Qtr4
2002                                     0.009444147
2003 0.015510690 0.021046915 0.025404011 0.029116231
2004 0.032405956 0.035391201 0.038143522 0.040710188
```

for log-data, (not original gnp data)

Example

- Forecasts from ARIMA(0,1,2)



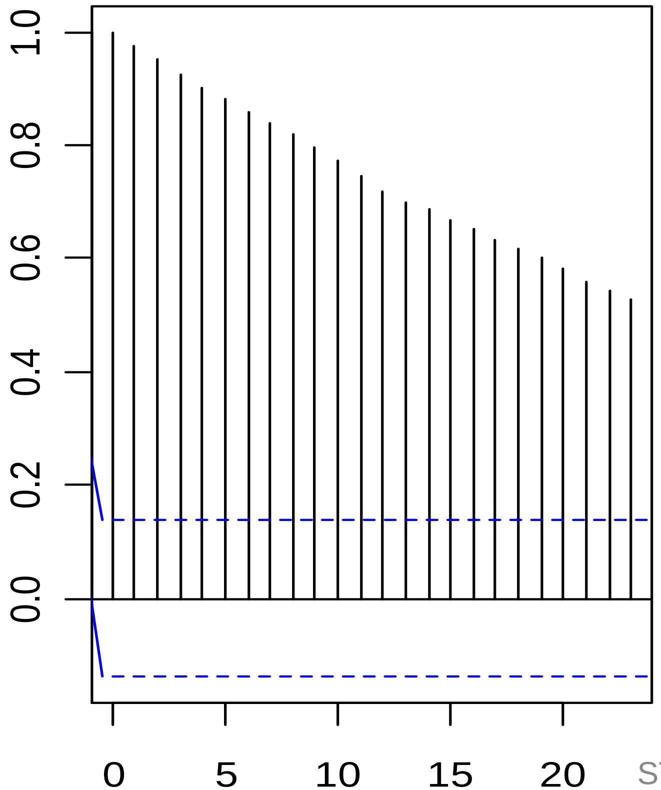
for original
gnp data

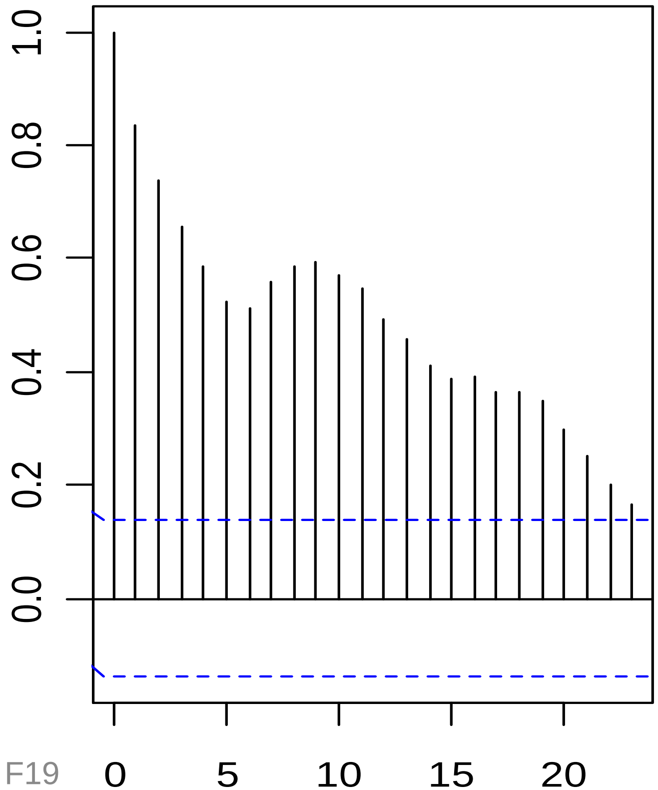
Unit Root Testing

- It is often difficult to distinguish Random Walk from AR(1) with $\phi \approx 1$ (especially for small n)
  - Using confidence interval for φ from AR(1) model $X_t = \varphi X_{t-1} + W_t$ has poor performance
- Unit root tests are designed to test H_0: $\varphi=1$ vs H_1: $|\varphi|<1$, for $X_t = \varphi X_{t-1} + W_t$ model
  - Most common test is the augmented Dickey-Fuller (ADF) test

Example

- ACF plot (n=200) for
  - Random Walk
  - AR(1) $X_t = .9 X_{t-1} + W_t$ .95





Example

- ADF test in R using `ADF.test()` function

load "*tseries*" package — data

```
library(tseries);   adf.test( x )
```

- Results:

for Random Walk (n=200)

```
Augmented Dickey-Fuller Test
Dickey-Fuller = -1.6749, Lag order = 5, p-value = 0.7123
alternative hypothesis: stationary

Augmented Dickey-Fuller Test
Dickey-Fuller = -3.922, Lag order = 5, p-value = 0.0142
alternative hypothesis: stationary
```

for AR(1) w/ φ=.9 (n=200)

Seasonality

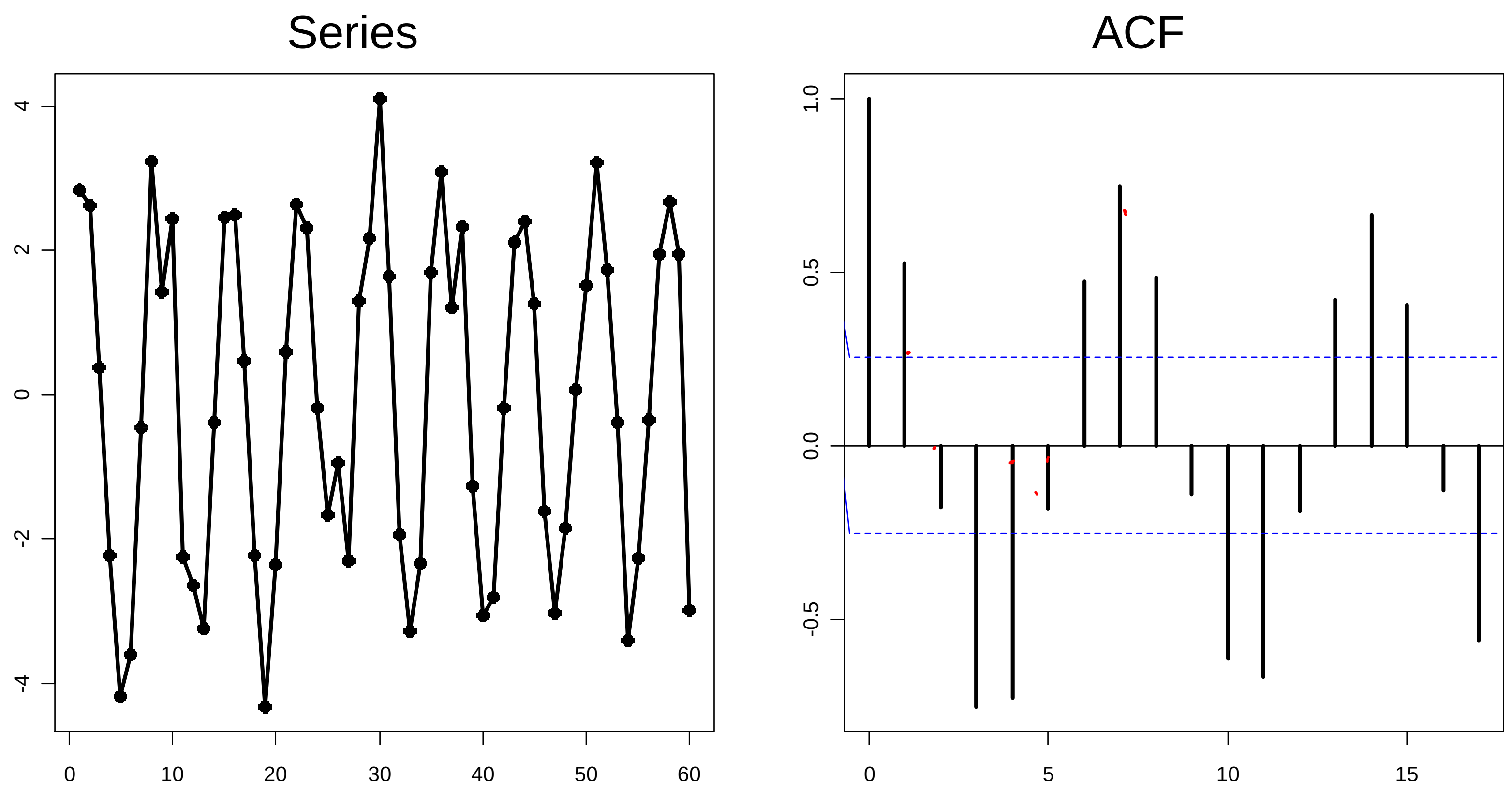
- Seasonality is *periodic pattern* in TS
  - E.g. Increase in summer sales of suntan lotions
- Period of seasonal pattern denoted by s
  - E.g. monthly data w/ annual pattern → s=12, daily data w/ weekly pattern → s=7, etc
- 2 ways to model seasonality, depending on its type:
  - Deterministic Seasonality
  - Stochastic Seasonality

Deterministic Seasonality

- Consider series $X_t = S(t) + Z_t$, where:
  - Z_t is stationary series
  - S(t) is deterministic seasonal component
    - i.e. periodic function w/ period s: $S(t) = S(t+s), \ \forall t$
- To estimate deterministic S(t), fit #s *separate means* (one for each time within the period)
  - E.g. $\hat{S}(t) = \begin{cases} \hat{\mu}_1, & \text{for } t = 1, 1+s, 1+2s, \ldots \\ \hat{\mu}_2, & \text{for } t = 2, 2+s, 2+2s, \ldots \\ \vdots & \\ \hat{\mu}_s, & \text{for } t = s, 2s, 3s, \ldots \end{cases}$ using ANOVA

Example

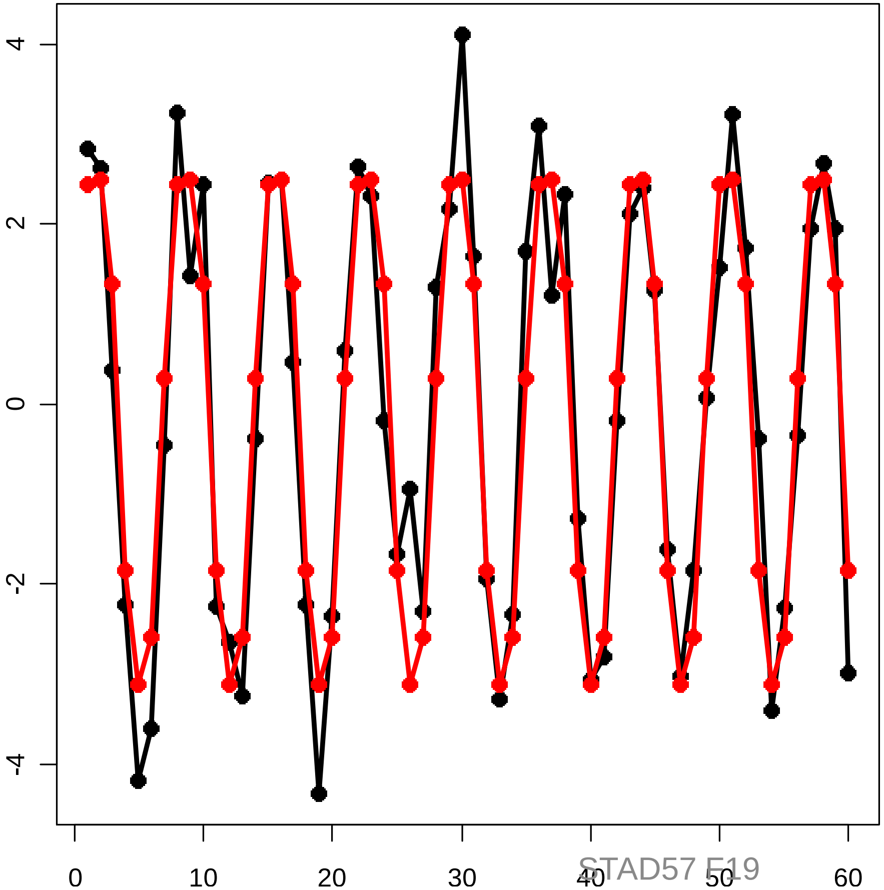
- Consider series with s=7



For seasonal series, ACF will typically exhibit pattern with same period (s)

Example

- Series with estimated $\hat{S}(t)$:



$$\hat{S}(t): \begin{cases} \hat{\mu}_1 = & 0.2893 \\ \hat{\mu}_2 = & 2.4413 \\ \hat{\mu}_3 = & 2.4938 \\ \hat{\mu}_4 = & 1.3351 \\ \hat{\mu}_5 = & -1.8420 \\ \hat{\mu}_6 = & -3.1090 \\ \hat{\mu}_7 = & -2.5936 \end{cases}$$

Further analysis (e.g. fit ARMA model) uses de-seasonalized series: $\hat{Z}_t = X_t - \hat{S}(t)$

Stochastic Seasonality

- More common and flexible situation is when seasonality is *stochastic*: $X_t = S_t + Z_t$, where:
  - S_t is RV, typically dependent on past of X_t
  - E.g. monthly sales w/ annual pattern, where this year's Jan sales depend on last year's Jan sales
- Simple way to model this behavior is to look at auto-regression at *multiple lags of period s*
  - E.g. For monthly data w/ annual pattern (s=12), can use: $X_t = \underbrace{\Phi X_{t-12}}_{S_t = \Phi X_{t-12}} + W_t$

$S_t = \Phi X_{t-12}$

Pure Seasonal ARMA Model

- More generally, a pure *seasonal ARMA* model of order (P,Q) and period s, denoted by *SARMA*(P,Q)$_s$, is given by:

$X_t = S_t + W_t$

$$X_t = \Phi_1 X_{t-s} + \Phi_2 X_{t-2s} + \ldots + \Phi_P X_{t-Ps} +$$

seasonal component is linear fn of lagged X's & W's

error

$$+ W_t + \Theta_1 W_{t-s} + \Theta_2 W_{t-2s} + \ldots + \Theta_Q W_{t-Qs} \Leftrightarrow$$

$$\Leftrightarrow \Phi(B^s) X_t = \Theta(B^s) W_t, \quad \text{where:}$$

seasonal $\mathrm{AR}(P)$ polynomial: $\Phi(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \ldots - \Phi_P B^{Ps}$

seasonal $\mathrm{MA}(Q)$ polynomial: $\Theta(B^s) = 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \ldots + \Theta_Q B^{Qs}$

Pure Seasonal ARMA Model

- The pure seasonal SARMA$(P,Q)_s$ model is causal / invertible if and only if the roots of $\Phi(z^s)$ / $\Theta(z^s)$ lie outside the unit circle:

$$\Phi(z^s) \neq 0 \,/\, \Theta(z^s) \neq 0, \forall\, |z| \leq 1$$

- ACF / PACF of pure seasonal SARMA$(P,Q)_s$ behaves similarly to ACF / PACF of usual ARMA(p,q) *at multiple lags of s*, and is 0 elsewhere

Example

- Find ACF of $X_t = \Phi X_{t-s} + W_t$ (SAR(1)$_s$)

$$X_t = \Phi X_{t-s} + W_t = \Phi^2 X_{t-2s} + \Phi W_{t-s} + W_t$$

$$= \Phi^3 X_{t-3s} + \Phi^2 W_{t-2s} + \Phi W_{t-s} + W_t$$

$$\vdots$$

$$= \sum_{j=0}^{\infty} \Phi^j W_{t-j\cdot s} \Rightarrow$$

$$\Rightarrow \gamma(0) = Var[X_t] = \sum_{j=0}^{\infty} (\Phi^j)^2 \cdot \underbrace{Var[W_{t-js}]}_{=\sigma_w^2} = \sigma_w^2 \cdot \frac{1}{1-\Phi^2}$$

$$\gamma(h) = Cov(X_{t+h}, X_t) = Cov\left(\sum_{j=0}^{\infty} \Phi^j \cdot W_{t-j\cdot s}, \sum_{k=0}^{\infty} \Phi^k \cdot W_{t+h-js}\right) =$$

$$= \begin{cases} 0, & \text{if } h \text{ is not multiple of } s \\ \Phi^n / 1-\Phi^2 & h = n\cdot s \end{cases}$$

Pure Seasonal ARMA Model

- Similarly, for SMA(1)$_s$ the ACF is:

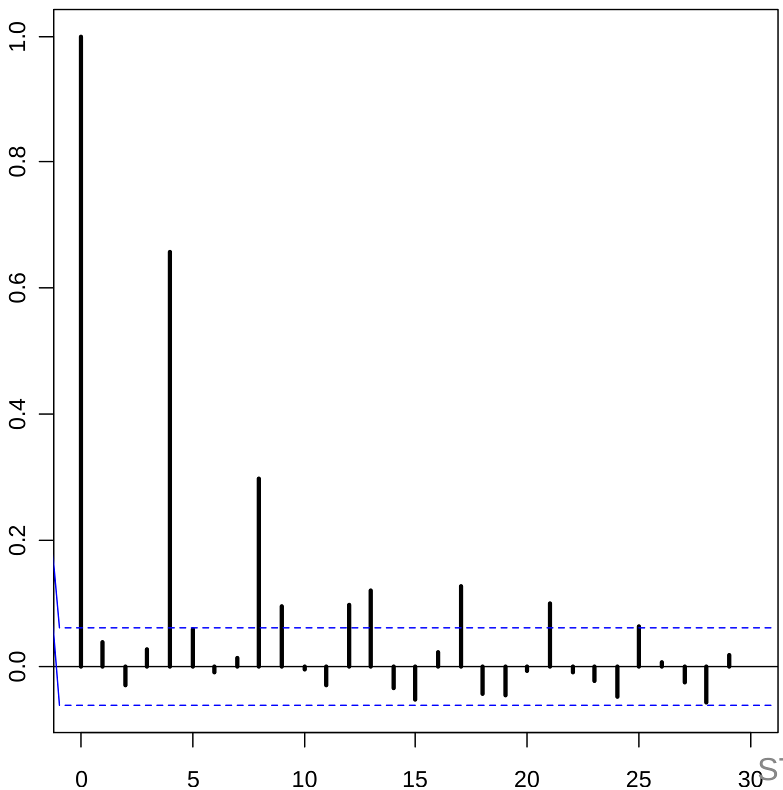
$$\rho(h)=\begin{cases} 1, & h=0 \\ \Theta/(1+\Theta^2), & h=s \\ 0, & \text{otherwise} \end{cases}$$

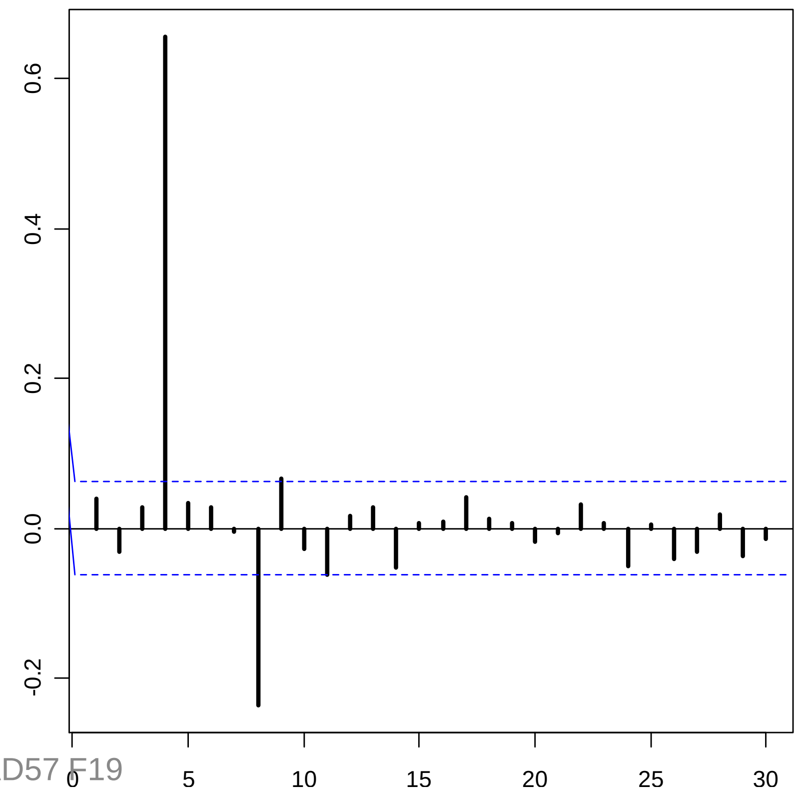
- Following table describes behavior of pure SARMA(P,Q)$_s$ models

	SAR(P)$_s$	**SMA(Q)$_s$**	**SARMA(P,Q)$_s$**
ACF	Tails off at lags k·s, k≥1	Cuts off at lag Q·s	Tails off at lags k·s, k≥1
PACF	Cuts off at lag P·s	Tails off at lags k·s, k≥1	Tails off at lags k·s, k≥1

Example

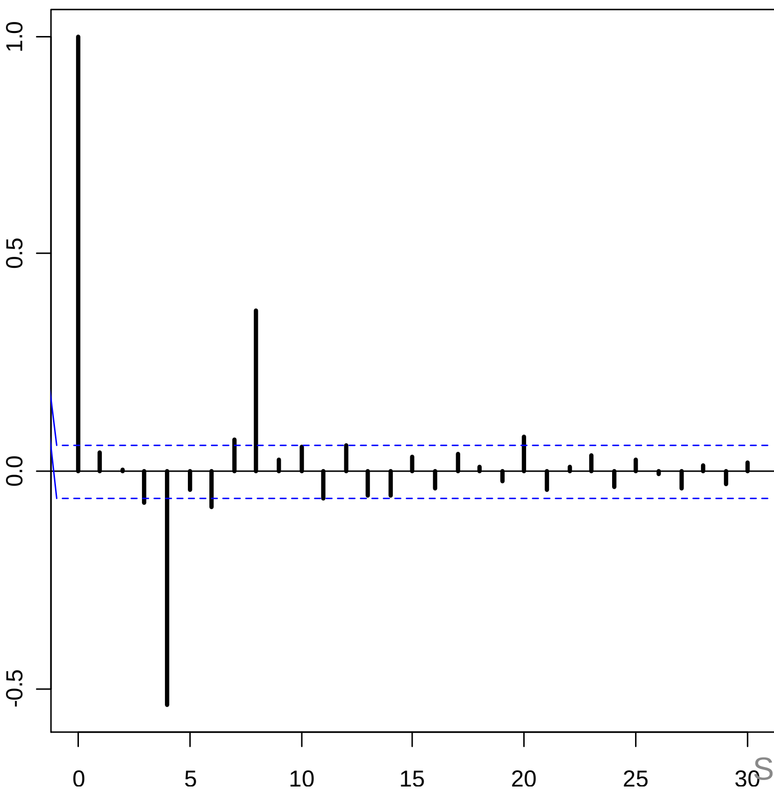
- SAR(2)$_4$ model:
  - ACF s = 4
  - PACF

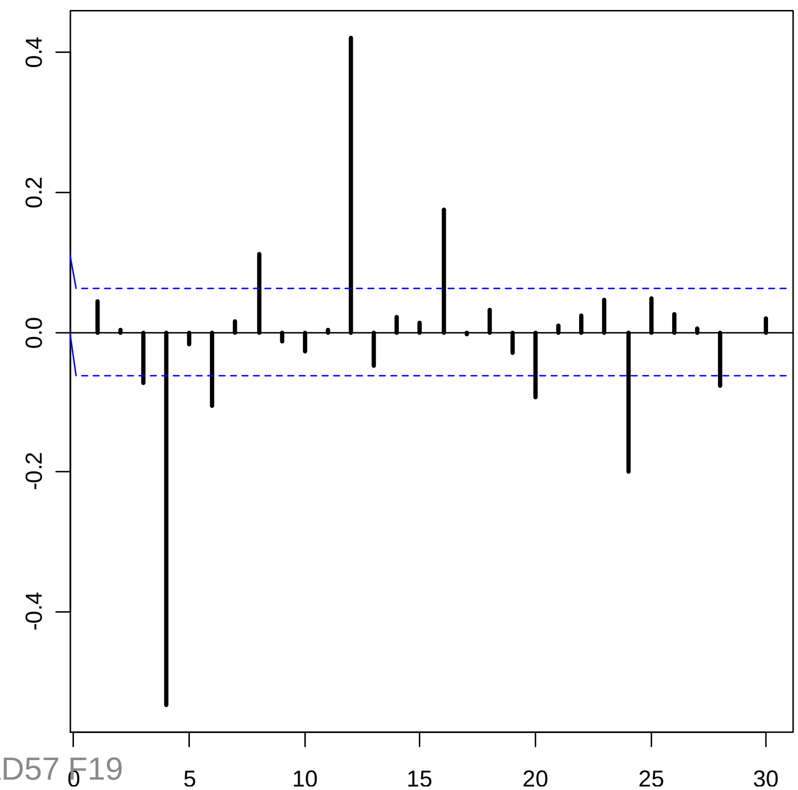




Example

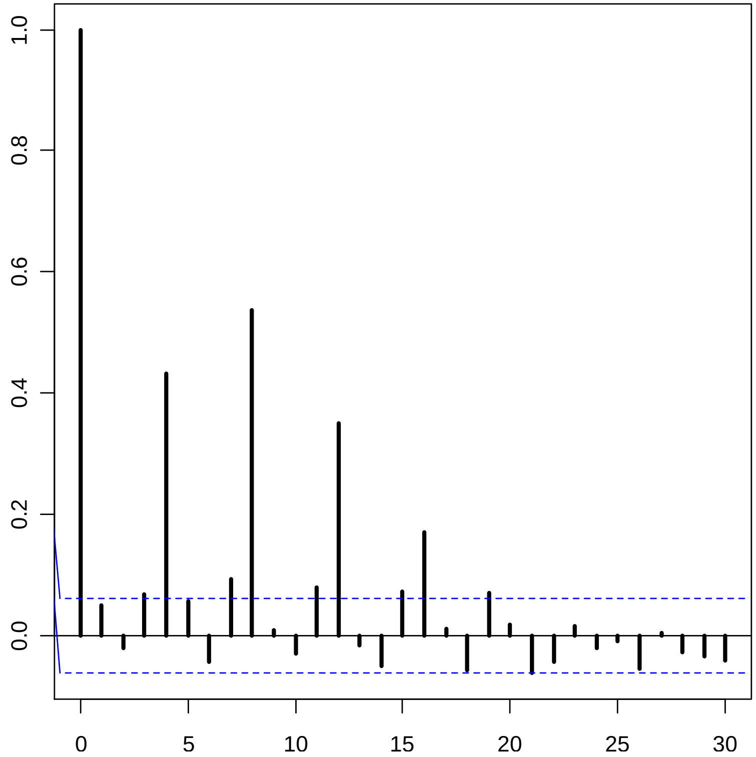
- SMA$(2)_4$ model:
  - ACF
  - PACF

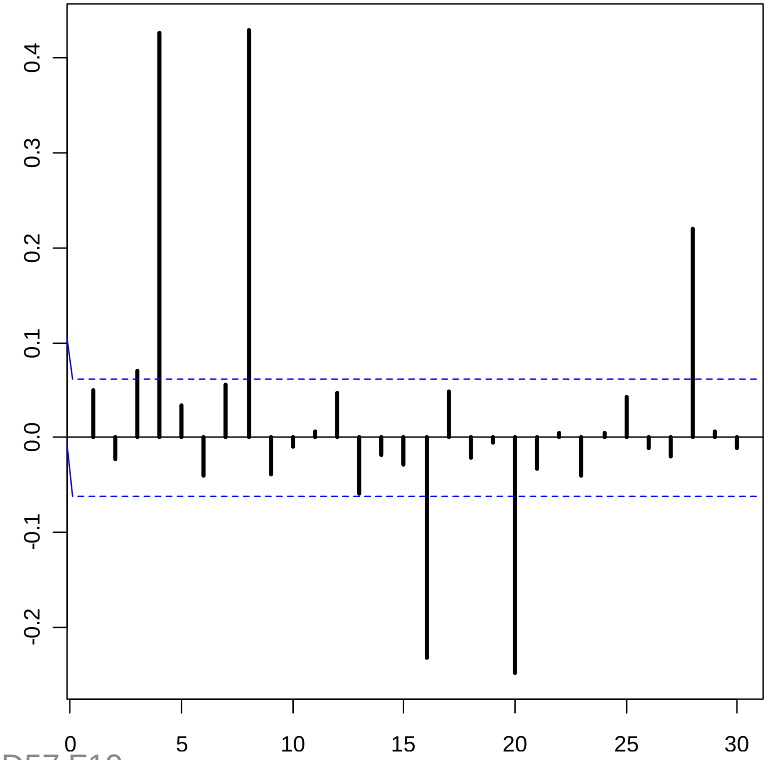




Example

- SARMA$(2,2)_4$ model:
  - ACF
  - PACF





Multiplicative Seasonal ARMA Model

- Can also combine seasonal SARMA(P,Q)$_s$ with simple ARMA(p,q) model, by multiplying their corresponding polynomials:

$$\Phi(B^s)\varphi(B)X_t = \Theta(B^s)\theta(\mathrm{B})W_t$$

  - Called *multiplicative* SARMA(p,q)×(P,Q)$_s$ model
  - E.g. monthly sales depend on last year's sales *and* last month's sales → use multiplicative SARMA(1,0)×(1,0)$_{12}$ model:

$$\Phi(B^{12})\varphi(B)X_t = W_t \Rightarrow (1-\Phi B^{12})(1-\varphi B)X_t = W_t \Rightarrow$$

$$\Rightarrow X_t = \varphi X_{t-1} + \Phi X_{t-12} - \varphi\Phi X_{t-13} + W_t$$

Example

ARMA(1,1)

$p=1, q=1$

seasonal

$P=1, Q=0$

- Write general form of SARMA(1,1)×(1,0)$_s$ model

$$\Phi(B^s)\varphi(B)X_t = \Theta(B^s)\cdot\theta(B)\cdot W_t \Rightarrow$$

$$(1-\Phi B^s)\cdot(1-\varphi B)\cdot X_t = (1+\Theta B^s)W_t \Rightarrow$$

$$(1-\varphi B-\Phi B^s+\varphi\Phi B^{s+1})X_t = (1+\Theta B^s)W_t \Rightarrow$$

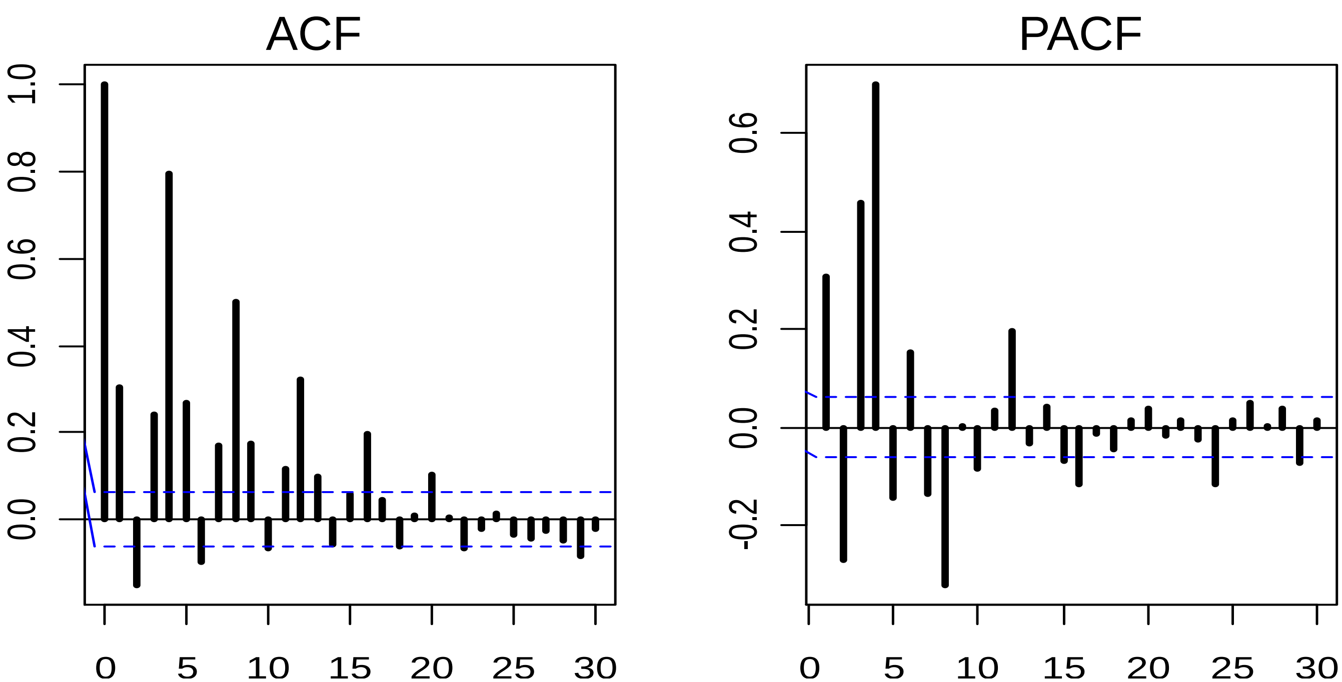
$$\Rightarrow X_t = \varphi X_{t-1}+\Phi X_{t-s}-\varphi\Phi X_{t-s-1}+W_t+\Theta W_{t-s}$$

Example

p q P Q

- Consider SARMA(1,1)×(1,1)$_4^{s}$ model:

$$(1-.6B^4)(1+.3B)X_t = (1+.6B^4)(1+.6B)W_t$$



Even for multiplicative SARMA models, ACF will typically have pattern with period (s)

Integrated Seasonal ARMA Model

- Often, seasonal component is not stationary
  - E.g. $X_t = S_t + W_t$, with $S_t = S_{t-s} + V_t$
  - In such cases, ACF at lags that are multiples of s does not tail off exponentially fast
- To make series stationary can use *seasonal differencing*, i.e. differencing at lag s
  - Define $\nabla_s^D = (1 - B^s)^D$ to be the D^th^-order difference operator *at lag* s
  - Applying the appropriate order seasonal differencing, $\nabla_s^D X_t$ will be stationary

Example

- If $X_t = S_t + W_t$, with $S_t = S_{t-s} + V_t$ and $W_t, V_t \sim WN$, show that $\nabla_s^1 X_t = \nabla_s X_t$ is stationary

$\nabla_s X_t = (1 - B^s) \cdot X_t = X_t - X_{t-s} =$

$= (S_t + W_t) - (S_{t-s} + W_{t-s}) =$

$= \underbrace{(S_t - S_{t-s})}_{= V_t} + (W_t - W_{t-s}) =$

$= V_t + \underbrace{W_t + W_{t-s}}_{\text{stationary}}$

Integrated Seasonal ARMA Model

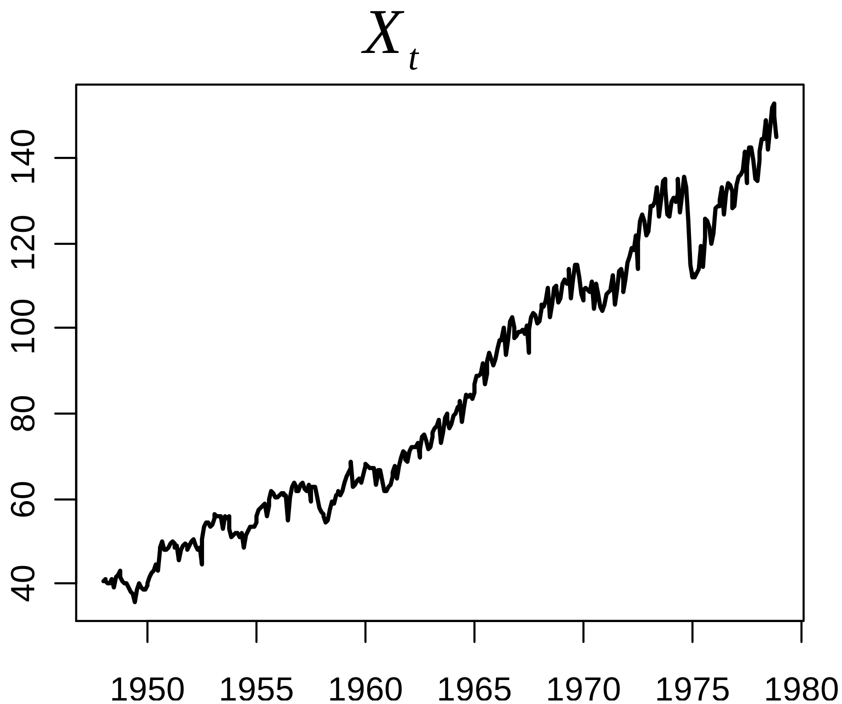
- The multiplicative seasonal autoregressive *integrated* moving average model, or SARIMA(p,d,q)×(P,D,Q)$_s$, is given by:

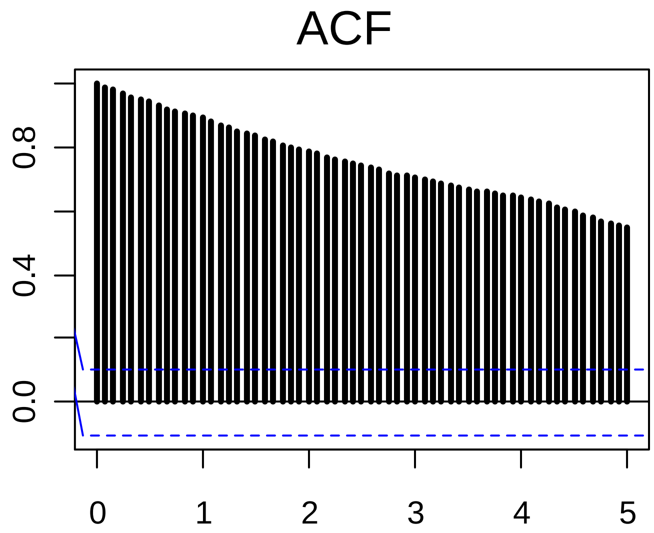
$$\Phi(B^s)\varphi(B)\nabla_s^D\nabla^d X_t = \Theta(B^s)\theta(\mathrm{B})W_t$$

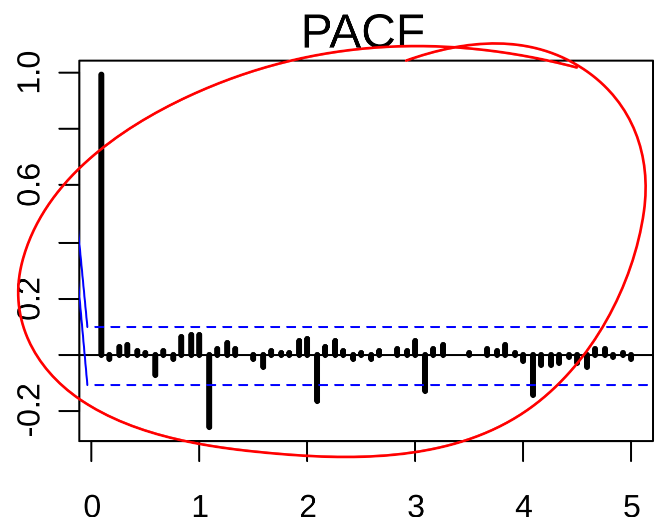
  - d is order of ordinary differencing, & D is order of seasonal differencing
  - Can still use AIC/BIC to select model order, and usual diagnostics to check model fit

Example

- US *monthly* production index:

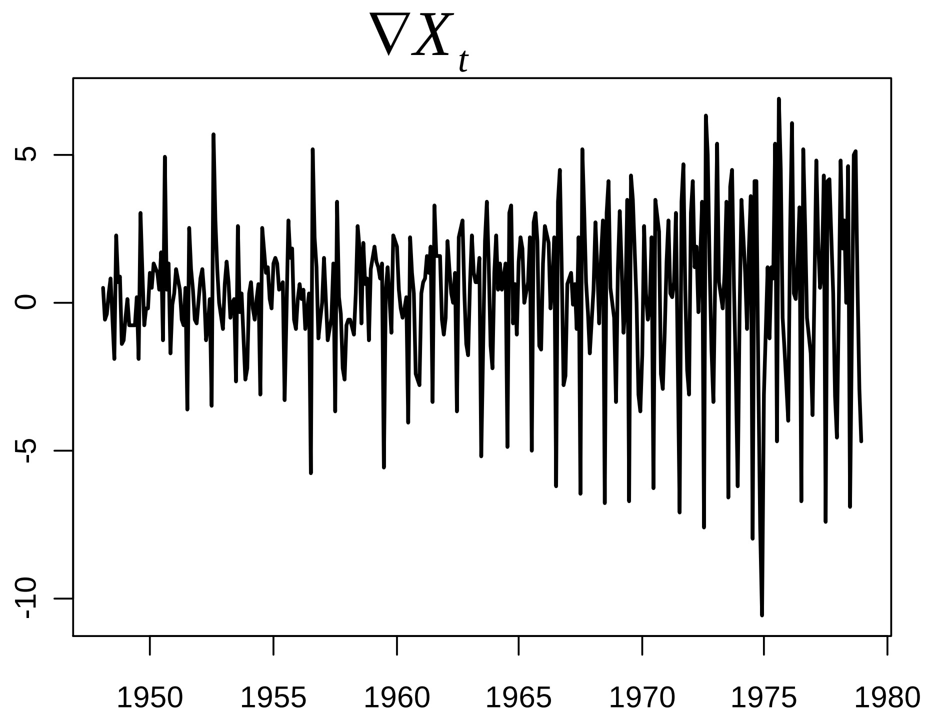


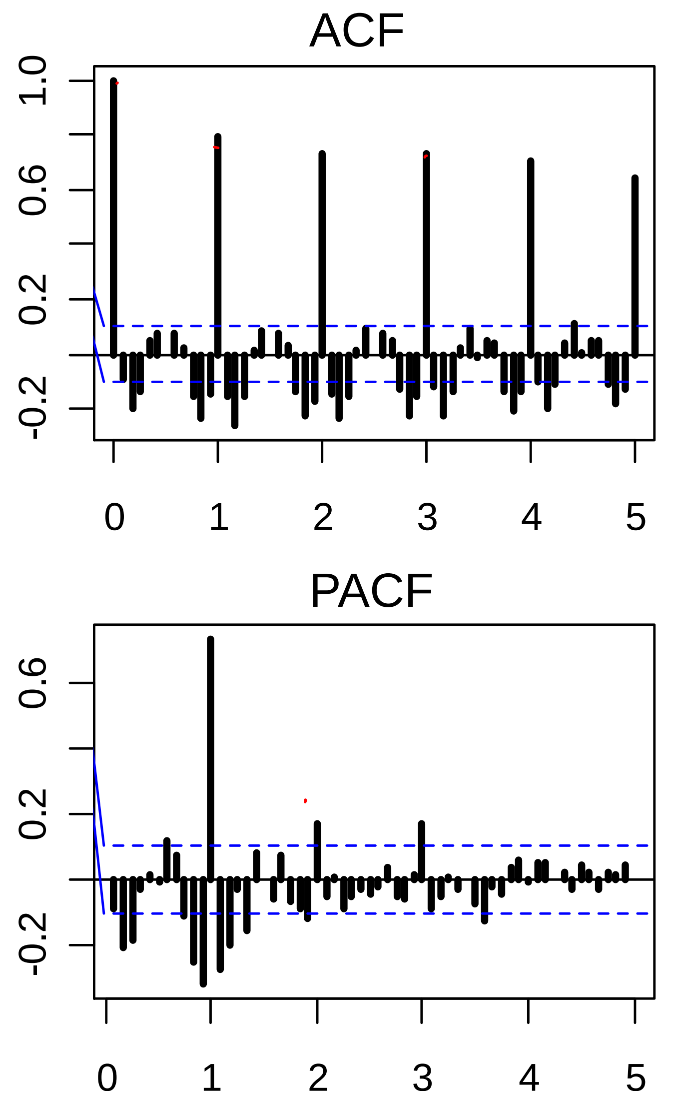




Example (cont'd)

- Differenced series:

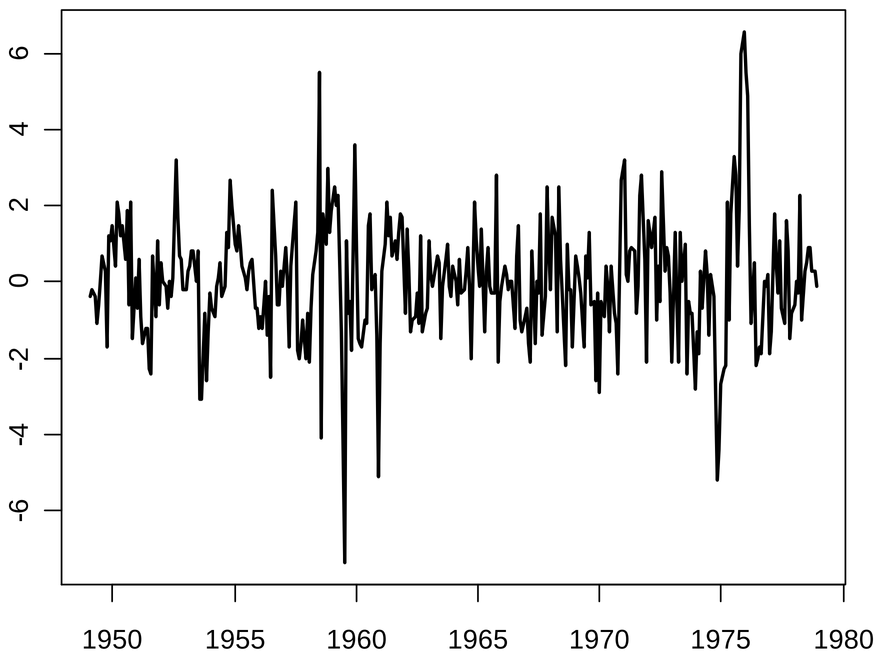


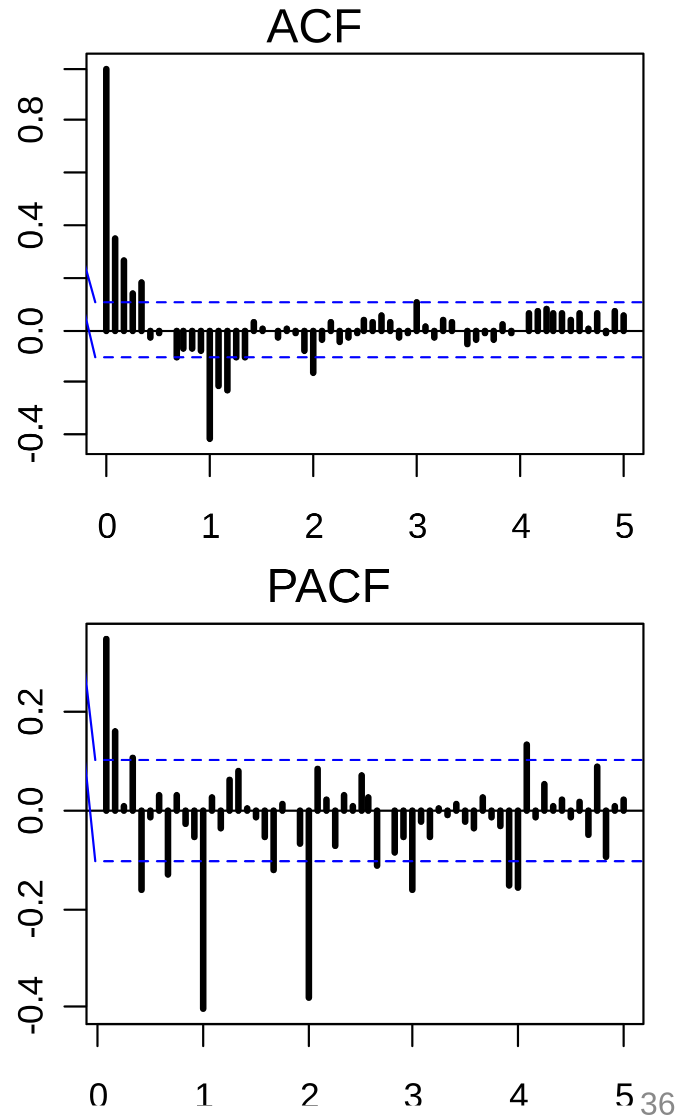


Example (cont'd)

- Seasonaly differenced (s=12) differenced series:

$\nabla_{12}\nabla X_t$





Example (cont'd)

seasonal

- Fitting SARIMA$(2,1,0)\times(0,1,3)_{12}$ in R

  p d q   P D Q

  (p,d,q)   (P,D,Q)   s

```
sarima(prodn, 2, 1, 0, 0, 1, 3, 12)
```

- Result:

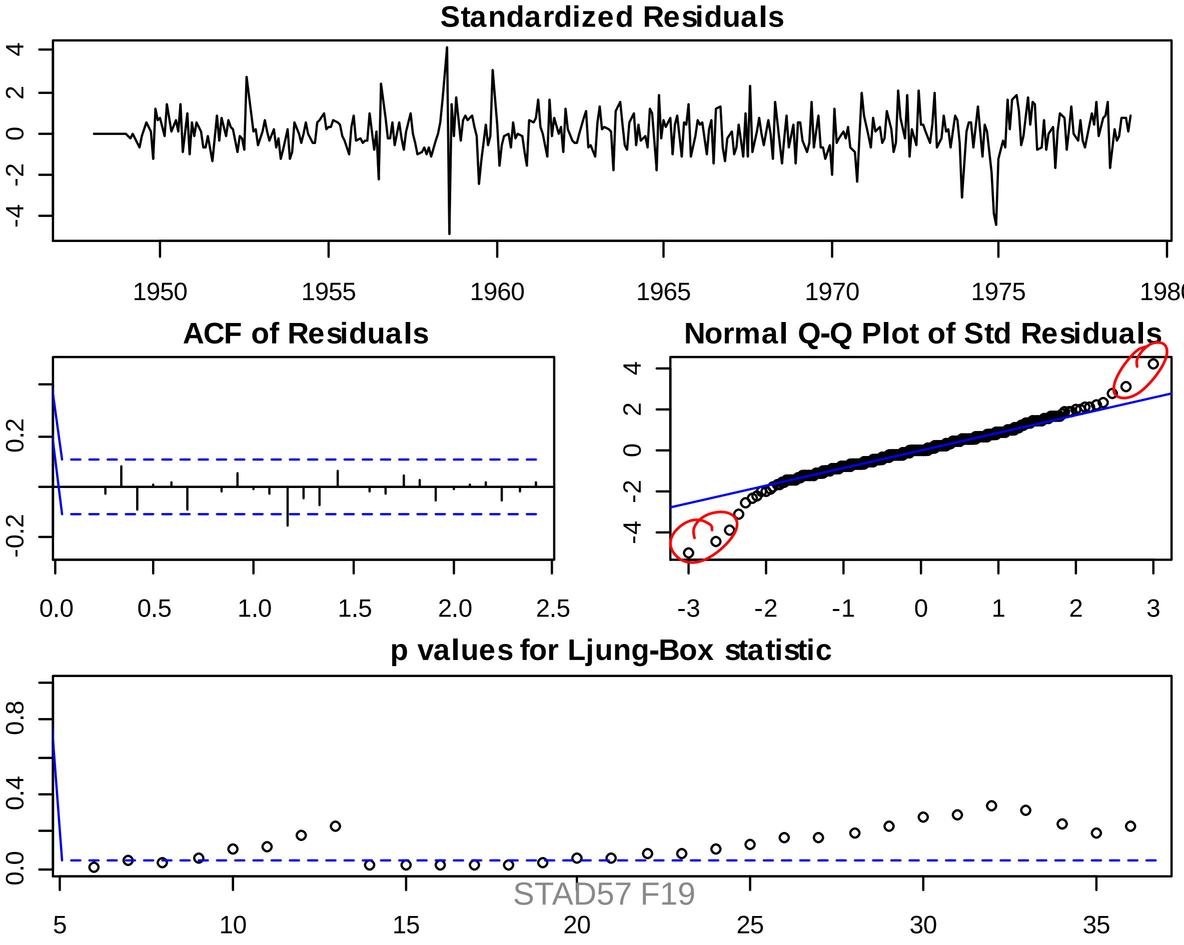
```
ARIMA(2,1,0)(0,1,3)[12]

Coefficients:
         ar1     ar2     sma1     sma2    sma3
      0.3038  0.1077  -0.7393  -0.1445  0.2815
s.e.  0.0526  0.0538   0.0539   0.0653  0.0526

sigma^2 estimated as 1.312:  log likelihood=-563.98
AIC=1139.97   AICc=1140.2   BIC=1163.26
```

ϕ_1 ϕ_2 Θ_1 Θ_2 Θ_3

fitted model: $(1-.30B-.11B^2)\nabla_{12}\nabla X_t = (1-.74B^{12}-.14B^{24}+.28B^{36})W_t$

Example (cont'd)

- Fit diagnostics

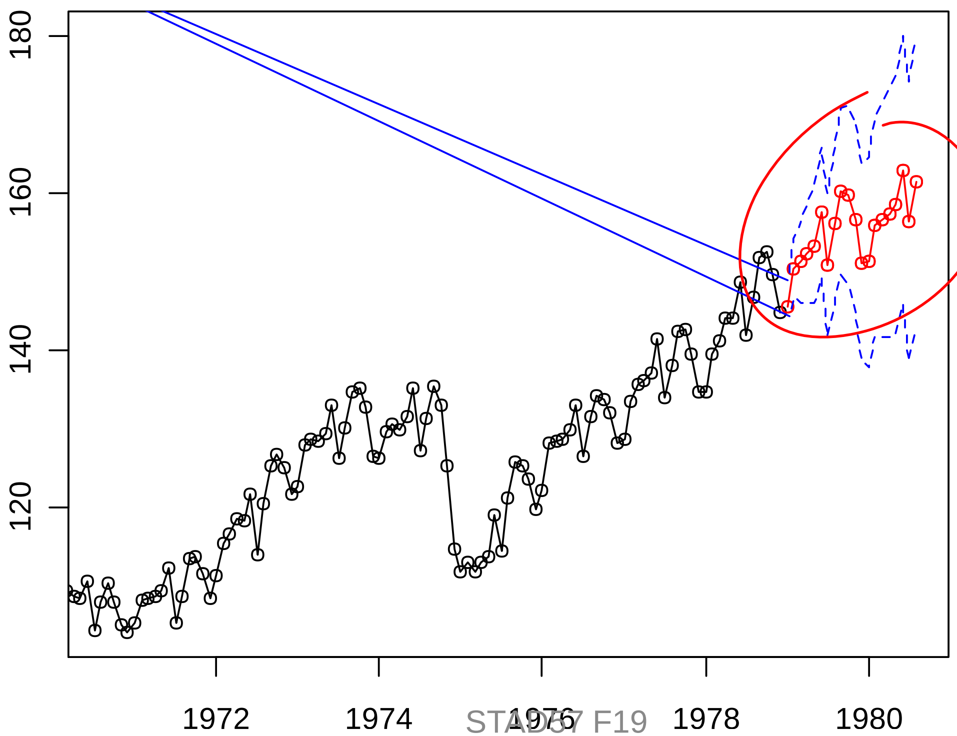


Example (cont'd)

- Model forecasts:

steps ahead (p,d,q) (P,D,Q) s

```
sarima.for(prodn, 20, 2, 1, 0, 0, 1, 3, 12)
```



for X_t, not $\nabla_{12} \nabla X$

SARIMA Model Selection

- SARIMA offers flexible framework for TS modeling & forecasting
  - Appropriate for TS w/ geometrically decaying ACF
- But model selection can be complicated
  - Have to compare AIC/BIC of all possible SARIMA$(p,d,q)\times(P,D,Q)_s$ combinations
- Can check ACF/PACF of (possibly differenced) data for guidance, and use *stepwise* selection method

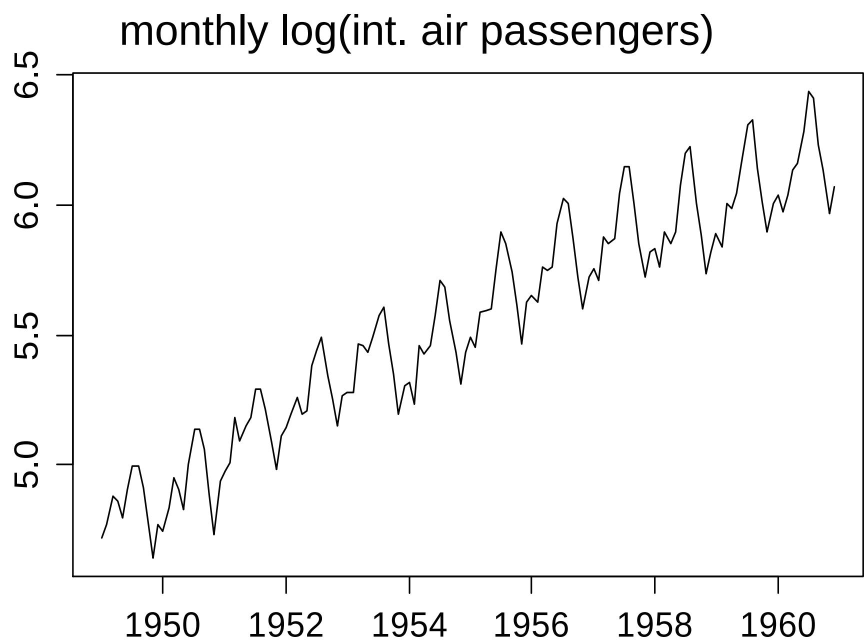
SARIMA Model Selection

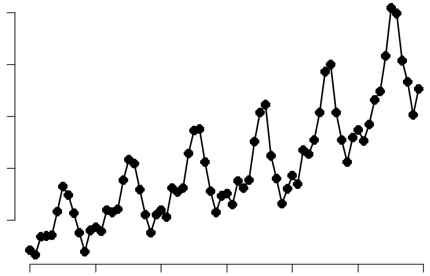
- For stepwise SARIMA model selection in R, use `auto.arima` function

data   selection criterion

```
auto.arima(lAP, ic="bic")
```

```
Series: lAP
ARIMA(0,1,1)(0,1,1)[12]

Coefficients:
          ma1     sma1
      -0.4018  -0.5569
s.e.   0.0896   0.0731

sigma^2 estimated as 0.001348:  log likelihood=244.7
AIC=-483.4   AICc=-483.21   BIC=-474.77
```





8. Multivariate Time Series

Univariate Time Series

- SARIMA works great for *causal* TS (i.e. $\rho(h) \sim \alpha^h$ as $h \to \infty$)
  - Addresses most univariate TS problems
- Cases where SARIMA fails & alternatives
  - Discrete-valued TS ⇨ Markov Chains
  - TS w/ *long memory*, i.e. $\rho(h) \sim 1/h$, ⇨ fractional integration
  - TS w/ *stochastic*, i.e. non-constant, variance ⇨(G)ARCH models

Multivariate Time Series

- When analyzing *multivariate* TS, there are many more interesting questions to ask:
  - Estimation & Model selection
  - Prediction (a.k.a. Forecasting)
  - Are different TS's related?
  - Does one TS *lead* the other(s)
  - How do changes in one affect other(s)

same as univariate

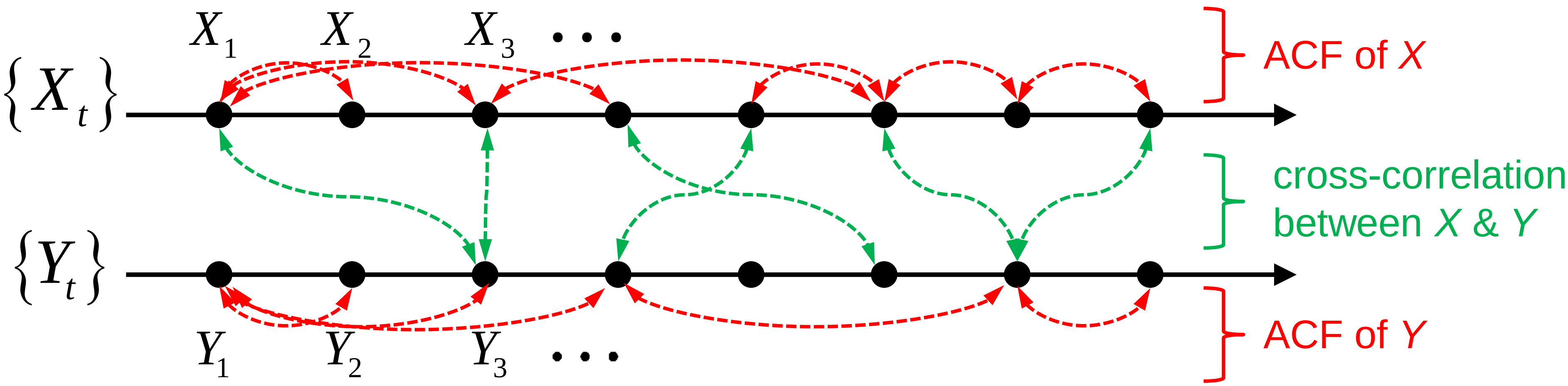
multivariate only

Multivariate Time Series

- Consider two TS $\{X_t\}, \{Y_t\}$. For *prediction* we can fit separate SARIMA models, so why bother with bivariate analysis?
  - If TS are independent / uncorrelated, then we can safely look at them separately
  - If TS are dependent, then predictions are better looked at jointly
- Cross-covariance looks at linear dependence of pairs of TS

Cross-Covariance

- Consider bivariate TS with components $\{X_t, Y_t\}$



  - To study dependence *within* X_t or Y_t , look at their respective autocovariances / ACFs
  - To study dependence *between* X_t *and* Y_t, look at *cross-covariance / cross-correlation* (CCF) function

Cross-Covariance / Cross-Correlation

- *Cross-covariance function* between X_t & Y_t

$$\gamma_{XY}(s,t) = Cov[X_s, Y_t] = E[(X_s - \mu_{X_s})(Y_t - \mu_{Y_t})]$$

- *Cross-correlation function* (*CCF*) b/t X_t & Y_t

$$\rho_{XY}(s,t) = \frac{\gamma_{XY}(s,t)}{\sqrt{\gamma_X(s,s)\gamma_Y(t,t)}}, \quad \left(\begin{array}{c}\text{where } \gamma_X(s,t) \text{ is} \\ \text{auto-cov. of } X_t\end{array}\right)$$

  - Note: $\gamma_{XY}(s,t) = \gamma_{YX}(t,s)$ & similarly for $\rho_{XY}(s,t)$
  - For trivariate TS $\{X_t, Y_t, Z_t\}$, would look at *all pair-wise* cross-covariances (γ_{XY}, γ_{XZ}, γ_{YZ}), & so on…

Joint Stationarity

- Two TS $\{X_t, Y_t\}$ are called *jointly stationary* if:
  1. X_t and Y_t are each stationary
  2. The cross-covariance is a function of $h = s - t$

$$\gamma_{XY}(h) = Cov[X_{t+h}, Y_t] = E[(X_{t+h} - \mu_X)(Y_t - \mu_Y)]$$

for $h = 0, \pm 1, \pm 2, \ldots$

- CCF of jointly stationary TS becomes

$$\rho_{XY}(h) = \frac{\gamma_{XY}(h)}{\sqrt{\gamma_X(0)\gamma_Y(0)}} = \frac{\gamma_{XY}(h)}{\sqrt{\sigma_X^2 \sigma_Y^2}}$$

Auto- vs Cross-Covariance

- For stationary $\{X_t\}$, autocov. is *symmetric*:

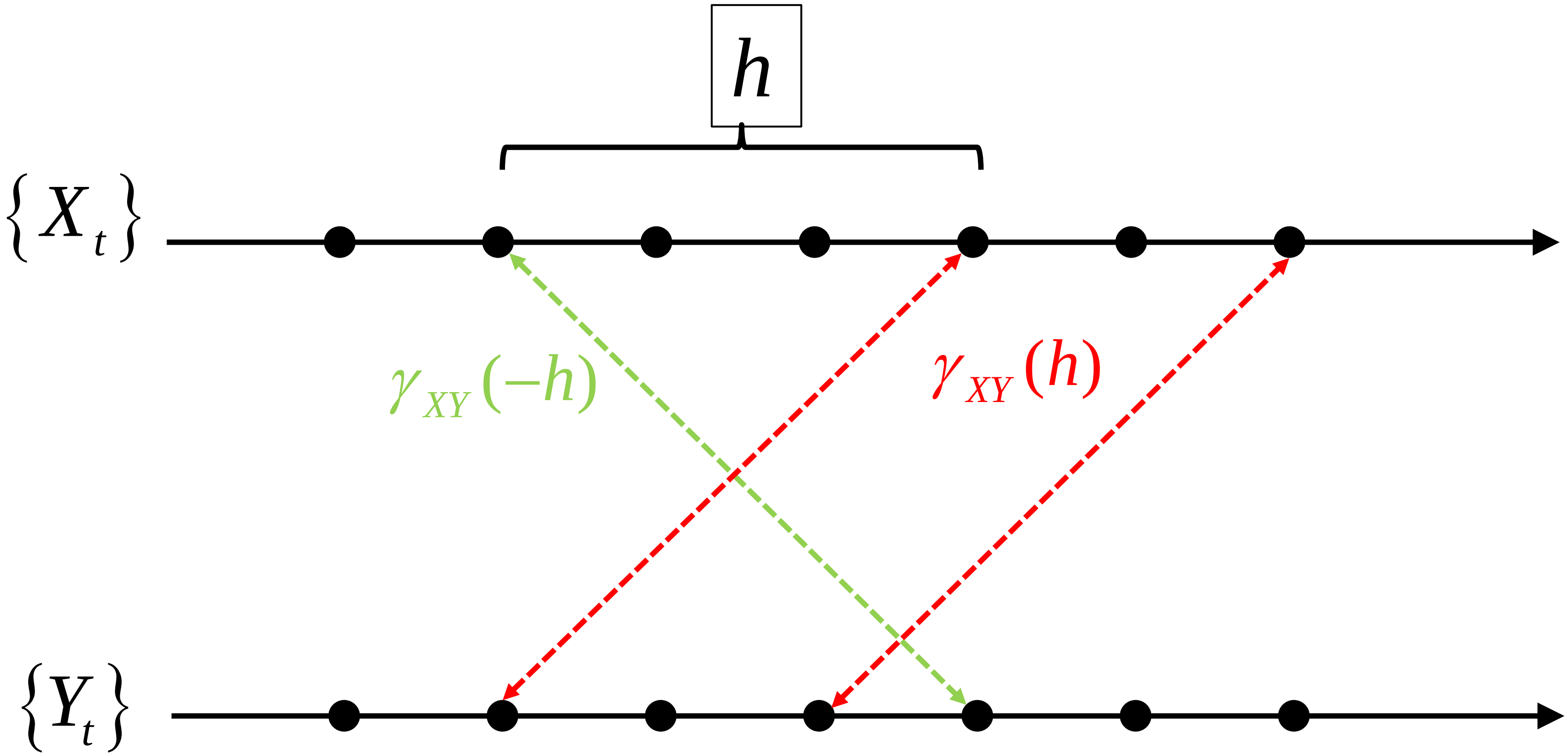
$$\gamma(h) = Cov[X_{t+h}, X_t] = Cov[X_t, X_{t+h}] = \gamma(-h)$$

  - That's why we only look at $\gamma(h)$ for $h \geq 0$
- For jointly stationary $\{X_t, Y_t\}$, cross-covariance is *NOT* symmetric

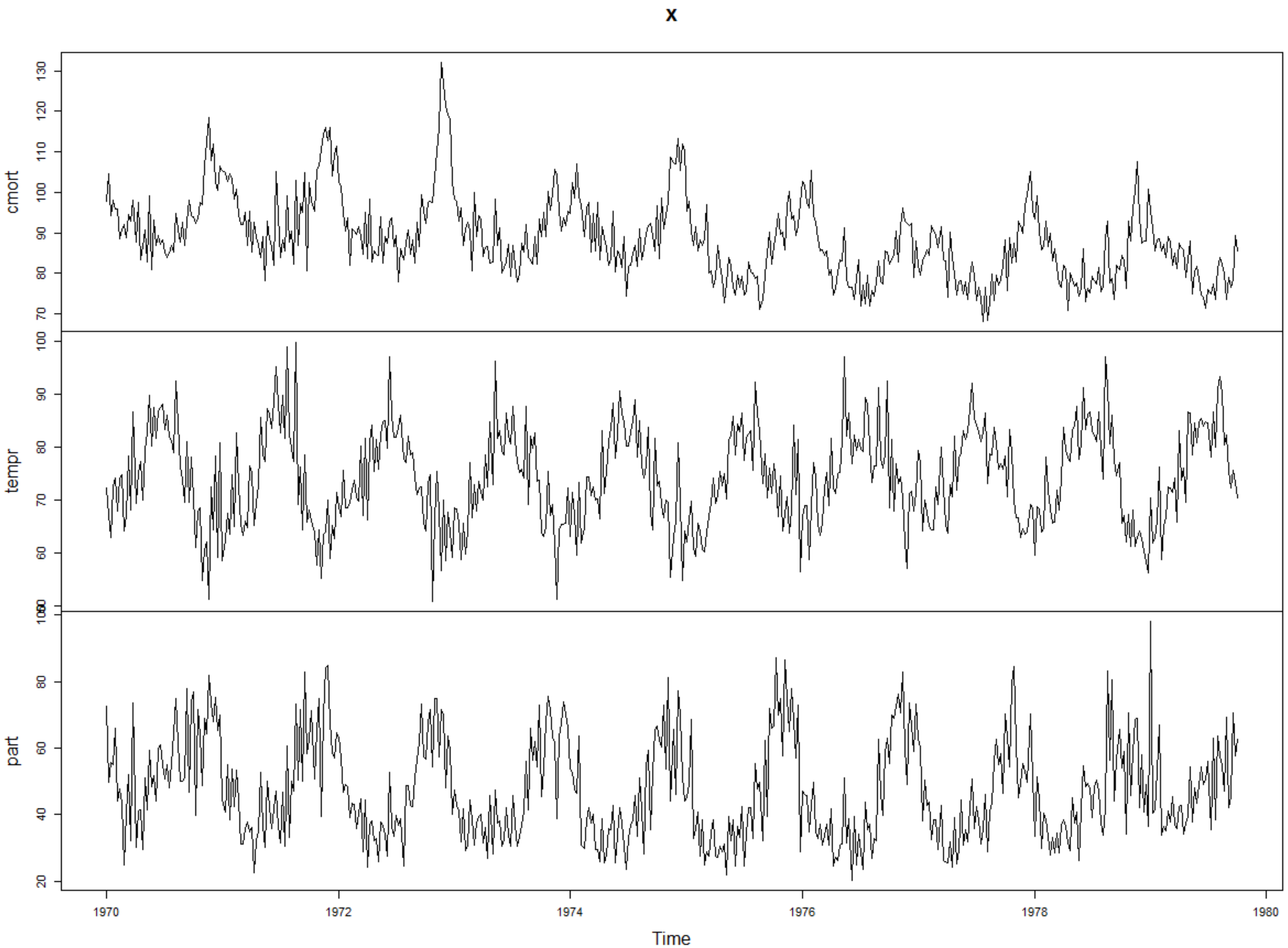
$$\gamma_{XY}(h) = \gamma_{XY}(t+h, t) = Cov[X_{t+h}, Y_t] \neq$$
$$\neq Cov[X_t, Y_{t+h}] = \gamma_{XY}(t-(t+h)) = \gamma_{XY}(-h)$$

  - That's why we need $\gamma_{XY}(h)$ for all h=0,±1,±2,…
  - However, $\gamma_{\mathbf{XY}}(h) = \gamma_{\mathbf{YX}}(-h)$

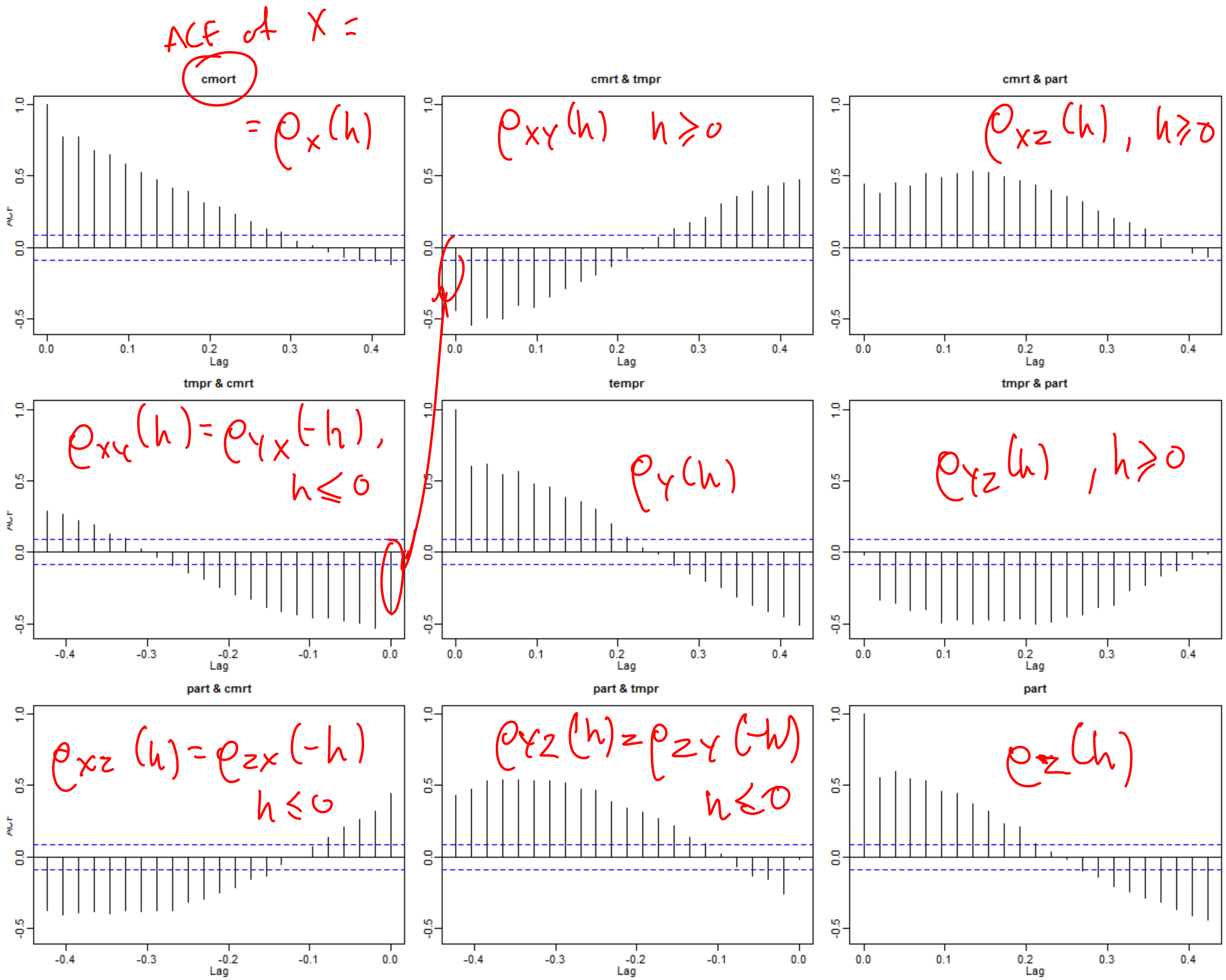
Auto- vs Cross-Covariance



Example



Example (Cont'd)



Example

- Consider 2D TS: $\begin{cases} X_t = W_t + W_{t-1} & \text{MA(1)} \\ Y_t = W_t - W_{t-1} & \text{MA(1)} \end{cases}$, where $W_t \sim WN(0, \sigma_W^2)$
  - Show that $\{X_t, Y_t\}$ is jointly stationary

First need to show that $\{X_t\}$ & $\{Y_t\}$ are individually stationary. But since they are both MA(1) $\Rightarrow$ they are both stationary with auto covariance function:

$$\left(\begin{aligned} \gamma_X(1) &= Cov(X_{t+1}, X_t) = Cov(W_{t+1} + W_t, W_t + W_{t-1}) = \\ &= \cancelto{0}{Cov(W_{t+1}, W_t)} + \cancelto{0}{Cov(W_{t+1}, W_{t-1})} + \underbrace{Cov(W_t, W_t)}_{=\sigma_W^2} + \\ &\quad + \cancelto{0}{Cov(W_t, W_{t-1})} \end{aligned} \right)$$

$$\gamma_X(h) = \begin{cases} 2\sigma_W^2, & h = 0 \\ \sigma_W^2, & h = 1 \\ 0, & h \geq 2 \end{cases}$$

& similarly: $\gamma_Y(h) = \begin{cases} 2\sigma_W^2, & h = 0 \\ -\sigma_W^2, & h = 1 \\ 0, & h \geq 2 \end{cases}$

Calculate cross-covariance to show joint stationarity

$$\gamma_{XY}(X_t, Y_t) = Cov(X_t, Y_t) = Cov(W_t + W_{t-1}, W_t - W_{t-1}) = Cov(W_t, W_t) - \underbrace{Cov(\cancel{W_t, W_{t-1}})}_{=0} + \underbrace{Cov(\cancel{W_{t-1}, W_t})}_{0} - Cov(W_{t-1}, W_{t-1}) = \sigma_w^2 - \sigma_w^2 = 0 = \gamma_{XY}(0)$$

$$\gamma_{XY}(X_{t+1}, Y_t) = Cov(W_{t+1} + W_t, W_t - W_{t-1}) = Cov(\cancel{W_{t+1}, W_t}) - Cov(\cancel{W_{t+1}, W_{t-1}}) + Cov(W_t, W_t) - Cov(\cancel{W_t, W_{t-1}}) = \sigma_w^2 = \gamma_{XY}(h=1)$$

$$\gamma_{XY}(X_{t-1}, Y_t) = Cov(W_{t-1} + W_{t-2}, W_t - W_{t-1}) = Cov(\cancel{W_{t-1}, W_t}) - Cov(W_{t-1}, W_{t-1}) + Cov(\cancel{W_{t-2}, W_t}) - Cov(\cancel{W_{t-2}, W_{t-1}}) = -\sigma_w^2 = \gamma_{XY}(h=-1)$$

$$\gamma_{XY}(X_{t+h}, Y_t) = \dots = 0, \quad \forall |h| \geq 2 \Rightarrow$$

$$\Rightarrow \gamma_{XY}(h) = \begin{cases} 0, & h = 0 \\ +\sigma_w^2, & h = 1 \\ -\sigma_w^2, & h = -1 \\ 0, & |h| \geq 2 \end{cases} \Rightarrow \text{Jointly stationary}$$

Example

$\{W_t\}$ uncorrelated w/ $\{X_t\}$

$\mu_X = 0$ & $\gamma_X(h)$ stationary

- Let $\{X_t\}$ be zero-mean, stationary TS & $W_t \sim WN(0, \sigma_w^2)$, and consider $Y_t = A \cdot X_{t-\ell} + W_t$
  - Show that $\{X_t, Y_t\}$ is jointly stationary

Know that $\{X_t\}$ is stationary ⇒ need to show that $\{Y_t\}$ is also stationary (i.e. μ_Y is constant & $\gamma_Y(h)$ is stationary).

$$\mu_Y = \mathbb{E}[Y_t] = \mathbb{E}[A \cdot X_{t-\ell} + W_t] = A \cdot \mathbb{E}[\underbrace{X_{t-\ell}}_{=0}] + \mathbb{E}[\underbrace{W_t}_{=0}] = 0$$

$$\gamma_Y(s,t) = Cov(Y_s, Y_t) = Cov(A \cdot X_{s-\ell} + W_s, A \cdot X_{t-\ell} + W_t) =$$

$$= A^2 \cdot Cov(X_{s-\ell}, X_{t-\ell}) + A \cdot \underbrace{Cov(X_{s-\ell}, W_t)}_{0} + A \cdot \underbrace{Cov(W_s, X_{t-\ell})}_{0} + Cov(W_s, W_t)$$

$$= A^2 \gamma_X(s - \ell - t + \ell) + A \cdot 0 + A \cdot 0 + \sigma_w^2 \cdot 1_{(s=t)} =$$

$$= A^2 \cdot \gamma_X(h) + 1_{(h=0)} \cdot \sigma_w^2, \text{ where } h = |s-t|$$

$\Rightarrow \{Y_t\}$ is stationary

To show joint stationarity:

$$\gamma_{XY}(\overset{s=t+h}{t+h}, t) = Cov(X_{t+h}, Y_t) =$$

$$= Cov(X_{t+h}, A \cdot X_{t-\ell} + W_t) =$$

$$= A \cdot Cov(X_{t+h}, X_{t-\ell}) + \underbrace{Cov(\cancel{X_{t+h}}, W_t)}_{0} =$$

$$= A\,\gamma_X(t+h-t+\ell) = A \cdot \gamma_X(h+\ell)$$

$\hookrightarrow$ function of $h = |s-t|$

Vector Auto-Regressive Model

- *Vector Auto-Regressive* (*VAR*) model

$$\mathbf{X}_t = \mathbf{\Phi}_1 \mathbf{X}_{t-1} + \cdots + \mathbf{\Phi}_p \mathbf{X}_{t-p} + \mathbf{W}_t$$

  - where

$$\mathbf{X}_t = \begin{bmatrix} X_{1,t} \\ \vdots \\ X_{k,t} \end{bmatrix} \& \, \mathbf{W}_t = \begin{bmatrix} W_{1,t} \\ \vdots \\ W_{k,t} \end{bmatrix}, \forall t \quad \mathbf{\Phi}_i = \begin{bmatrix} \varphi_{i:1,1} & \vdots & \varphi_{i:1,k} \\ \cdots & \ddots & \cdots \\ \varphi_{i:k,1} & \vdots & \varphi_{i:k,k} \end{bmatrix}, \forall i = 1, \ldots, p$$

$$\mathrm{Var}(\mathbf{W}_t) = \mathrm{Cov}(\mathbf{W}_t, \mathbf{W}_t) = \mathbf{\Sigma}_W = \begin{bmatrix} \sigma_{1,1}^2 & \vdots & \sigma_{1,k}^2 \\ \cdots & \ddots & \cdots \\ \sigma_{k,1}^2 & \vdots & \sigma_{k,k}^2 \end{bmatrix} \; \& \; \mathrm{Cov}(\mathbf{W}_t, \mathbf{W}_s) = \mathbf{0}, \forall s \neq t$$

Example

- 2D VAR(1) model:

$$\mathbf{X}_t = \mathbf{\Phi}_1 \mathbf{X}_{t-1} + \mathbf{W}_t \Leftrightarrow \begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} = \begin{bmatrix} \varphi_{1;1,1} & \varphi_{1;1,2} \\ \varphi_{1;2,1} & \varphi_{1;2,2} \end{bmatrix} \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \begin{bmatrix} W_{1,t} \\ W_{2,t} \end{bmatrix} \Leftrightarrow$$

$$\Leftrightarrow \begin{Bmatrix} X_{1,t} = \varphi_{1;1,1} X_{1,t-1} + \varphi_{1;1,2} X_{2,t-1} + W_{1,t} \\ X_{2,t} = \varphi_{1;2,1} X_{1,t-1} + \varphi_{1;2,2} X_{2,t-1} + W_{2,t} \end{Bmatrix}$$

- where $\begin{cases} \mathbb{E}\left[W_{1,t}\right] = \mathbb{E}\left[W_{2,t}\right] = 0, \\ \mathbb{V}\left[W_{1,t}\right] = \sigma_{1,1}, \ \mathbb{V}\left[W_{2,t}\right] = \sigma_{2,2}, \\ \mathrm{Cov}(W_{1,t}, W_{2,t}) = \sigma_{1,2} = \sigma_{2,1} \end{cases}$

Example

- Find CCF of 2D $\{\mathbf{W}_t\} \sim \text{WN}(\mathbf{0}, \mathbf{\Sigma}_W)$

$$\gamma_{W_1,W_2}(h) = Cov(W_{1,t+h}, W_{2,t}) = \begin{cases} \sigma_{1,2} = \sigma_{2,1}, & h = 0 \\ 0, & h \neq 0 \end{cases}$$

$$\rho_{W_1,W_2}(h) = \frac{\gamma_{W_1,W_2}(h)}{\sqrt{\gamma_{W_1}(0) \cdot \gamma_{W_2}(0)}} = \begin{cases} \dfrac{\sigma_{1,2}}{\sqrt{\sigma_{1,1} \cdot \sigma_{2,2}}}, & h = 0 \\ 0, & h \neq 0 \end{cases}$$

Example

- Fit VAR model with function `vars::VAR()`

```
VAR Estimation Results:
=======================

Estimated coefficients for equation cmort:
==========================================
Call:
cmort = cmort.l1 + tempr.l1 + part.l1 + const

   cmort.l1    tempr.l1     part.l1       const
 0.60149346 -0.30946101  0.07096225 54.94579126


Estimated coefficients for equation tempr:
==========================================
Call:
tempr = cmort.l1 + tempr.l1 + part.l1 + const

  cmort.l1   tempr.l1    part.l1      const
-0.1787076  0.5111817 -0.1412636 58.8456142


Estimated coefficients for equation part:
=========================================
Call:
part = cmort.l1 + tempr.l1 + part.l1 + const

   cmort.l1    tempr.l1     part.l1       const
-0.08082151 -0.45998718  0.57215788 61.58412402
```

→ VAR(1)

$\Phi = \begin{bmatrix} \phi_{11} & \phi_{12} & \phi_{13} \\ & & \\ & & \end{bmatrix}$

VAR Model

- Consider VAR(p) model

$$\mathbf{X}_t = \mathbf{\Phi}_1 \mathbf{X}_{t-1} + \cdots + \mathbf{\Phi}_p \mathbf{X}_{t-p} + \mathbf{W}_t, \ \{\mathbf{W}_t\} \sim \mathrm{WN}(\mathbf{0}, \mathbf{\Sigma}_W)$$

- Model can be written as Wold process

$$\mathbf{X}_t = \mathbf{W}_t + \mathbf{\Psi}_1 \mathbf{W}_{t-1} + \mathbf{\Psi}_2 \mathbf{W}_{t-2} + \cdots = \sum_{j=0}^{\infty} \mathbf{\Psi}_j \mathbf{W}_{t-j}$$

  - where $\mathbf{\Psi}$-matrices satisfy:

$$\mathbf{\Psi}_k = \sum_{j=0}^{\min(k,p)} \mathbf{\Psi}_{k-j} \mathbf{\Phi}_j \ \& \ \mathbf{\Psi}_0 = \mathbf{I}$$

Example

- Find Wold representation of VAR(1) model

$$\mathbf{X}_t = \mathbf{\Phi}\mathbf{X}_{t-1} + \mathbf{W}_t, \ \{\mathbf{W}_t\} \sim \mathrm{WN}(\mathbf{0}, \mathbf{\Sigma}_W)$$

$$\underline{X}_t = \underline{\Phi}\,\underline{X}_{t-1} + \underline{W}_t$$

$$= \underline{\Phi}\cdot(\underline{\Phi}\cdot X_{t-2} + \underline{W}_{t-1}) + \underline{W}_t$$

$$= \underline{\Phi}^2 X_{t-2} + \underline{\Phi}\,\underline{W}_{t-1} + W_t$$

$$= \underline{\Phi}^2\cdot(\underline{\Phi}\,X_{t-3} + \underline{W}_{t-2}) + \underline{\Phi}\,\underline{W}_{t-1} + W_t$$

$$\vdots$$

$$= \sum_{j=0}^{\infty} \underline{\Phi}^j\cdot \underline{W}_{t-j}, \quad \text{where } \underline{\Phi}^0 = \underline{I}$$

Example (cont'd)

- Find stationary variance-covariance matrix of VAR(1) model

$$\mathbb{E}[\underline{X}_t] = \mathbb{E}\left[\sum_{j=0}^{\infty} \underline{\underline{\Phi}}^j \cdot \underline{W}_{t-j}\right] = \sum_{j=0}^{\infty} \underline{\underline{\Phi}}^j \underbrace{\mathbb{E}[\underline{W}_{t-j}]}_{=\underline{0}} = \underline{0}$$

$$\mathrm{Var}[\underline{X}_t] = \mathrm{Var}\left[\sum_{j=0}^{\infty} \underline{\underline{\Phi}}^j \cdot \underline{W}_{t-j}\right] = \sum_{j=0}^{\infty} \underline{\underline{\Phi}}^j \cdot \underbrace{\mathrm{Var}[\underline{W}_{t-j}]}_{=\Sigma_w} \left(\underline{\underline{\Phi}}^j\right)^T =$$

$$= \sum_{j=0}^{\infty} \underline{\underline{\Phi}}^j \, \underline{\underline{\Sigma}}_w \cdot \left(\underline{\underline{\Phi}}^j\right)^T$$

VAR Model

- Any VAR(p) can be expressed as special VAR(1) model:

$$\mathbf{X}_t = \mathbf{\Phi}_1 \mathbf{X}_{t-1} + \cdots + \mathbf{\Phi}_p \mathbf{X}_{t-p} + \mathbf{W}_t \Leftrightarrow$$

$$\Leftrightarrow \begin{bmatrix} \mathbf{X}_t \\ \vdots \\ \mathbf{X}_{t-p+1} \end{bmatrix} = \begin{bmatrix} \mathbf{\Phi}_1 & \mathbf{\Phi}_2 & \cdots & \mathbf{\Phi}_p \\ \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{X}_{t-1} \\ \vdots \\ \mathbf{X}_{t-p} \end{bmatrix} + \begin{bmatrix} \mathbf{W}_t \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}$$

VAR Model

- A VAR(p) model is causal (stationary) if

$$\det\left(\mathbf{I} - \mathbf{\Phi}_1 z - \cdots - \mathbf{\Phi}_p z^p\right) \neq 0, \;\; \forall \, |z| \leq 1$$

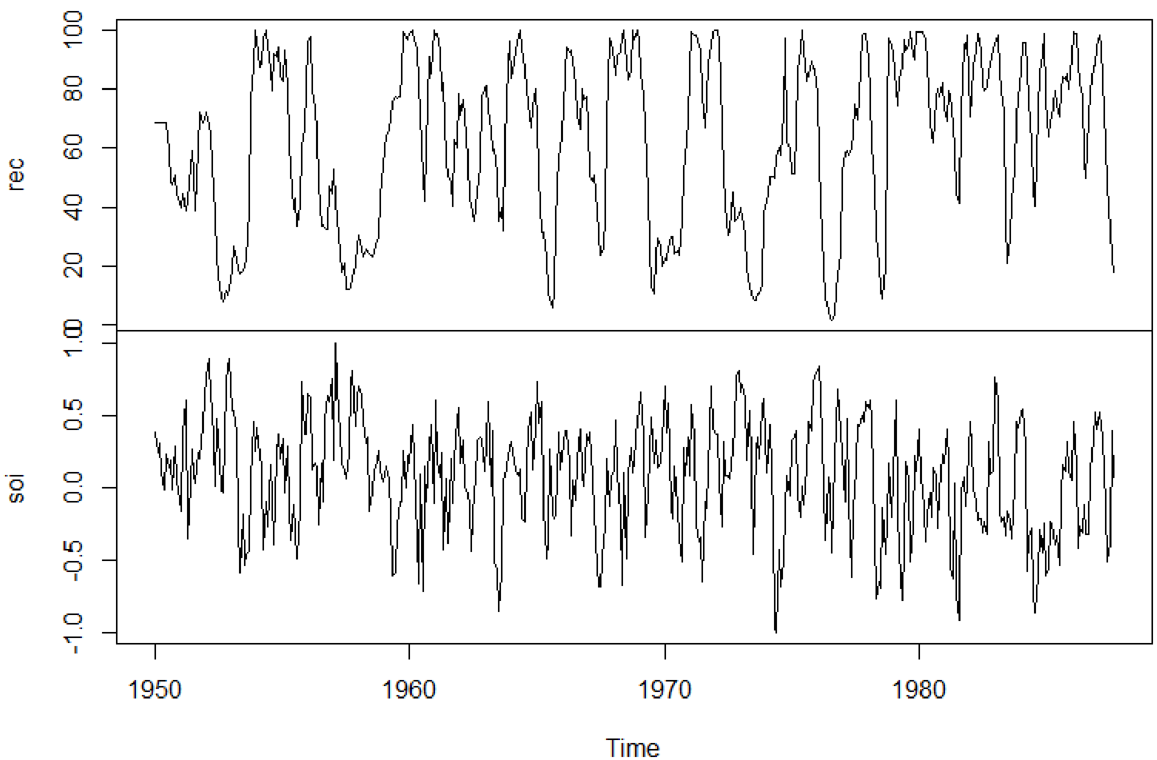
  - For VAR(1) ⇔ eigen-values of $\mathbf{\Phi}$ are all <1
  - Generally ⇔ eigen-values of $\begin{bmatrix} \mathbf{\Phi}_1 & \mathbf{\Phi}_2 & \cdots & \mathbf{\Phi}_p \\ \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \mathbf{0} \end{bmatrix}$ are all <1

VAR Estimation

- Function `VAR()` in package `vars` fits VAR(p) model
  - E.g. `VAR(X, p=2)`
- Model selection using AIC/BIC through `VARselect()` function
  - E.g. `VARselect(X, lag.max=15)`
  - Returns “AIC”, “BIC” & other criteria

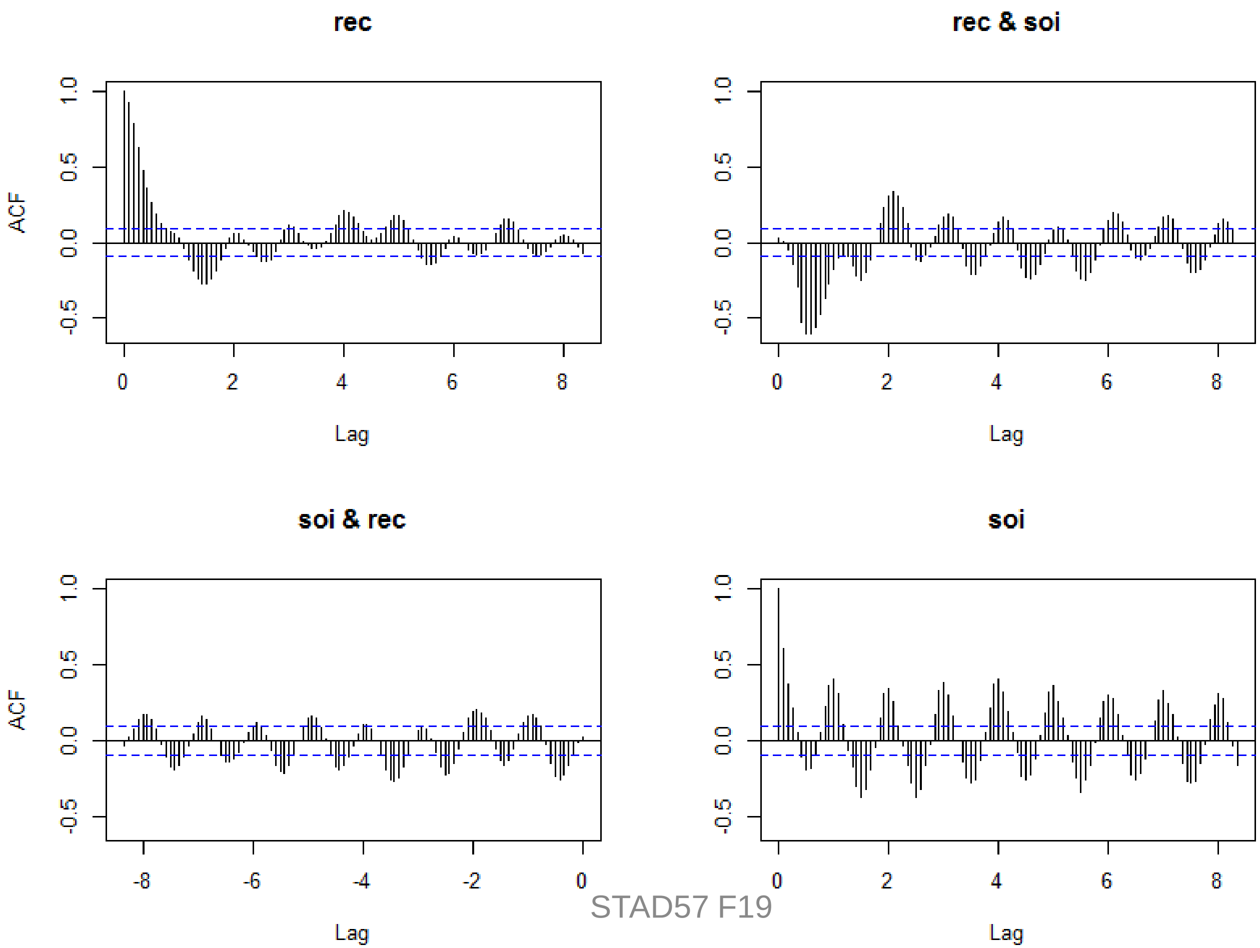
Example

- Monthly data: Southern Oscillation Index (soi) & Pacific Ocean # fish (rec)



Example (cont'd)

- ACF / CCF



Example (cont'd)

- VAR model selection

BIC = SC = "Schwartz Criterion"

```
> VARselect(X,lag.max=20)
$selection
AIC(n)  HQ(n)  SC(n) FPE(n)
    15      8      7     15

$criteria
                1        2        3        4        5        6        7        8        9       10
AIC(n)   2.402970 2.140677 2.112579 2.027653 1.650150 1.542299 1.488852 1.470325 1.464954 1.449333
HQ(n)    2.425237 2.177789 2.164536 2.094455 1.731797 1.638791 1.600188 1.596507 1.605981 1.605205
SC(n)    2.459377 2.234689 2.244197 2.196875 1.856978 1.786731 1.770889 1.789968 1.822202 1.844186
FPE(n) 11.055963 8.505208 8.269589 7.596325 5.207876 4.675494 4.432249 4.351002 4.327834 4.260922
              11       12       13       14       15       16       17       18       19       20
AIC(n) 1.444060 1.446837 1.445788 1.408018 1.402028 1.411512 1.407134 1.414192 1.424015 1.439889
HQ(n)  1.614776 1.632399 1.646194 1.623269 1.632124 1.656453 1.666920 1.688823 1.713491 1.744209
SC(n)  1.876517 1.916900 1.953455 1.953290 1.984905 2.031994 2.065221 2.109884 2.157312 2.210791
FPE(n) 4.238715 4.250746 4.246571 4.089489 4.065427 4.104585 4.087123 4.116601 4.157834 4.225032
```

- Can include deterministic seasonality w/ option `VARselect(…,season=s)`

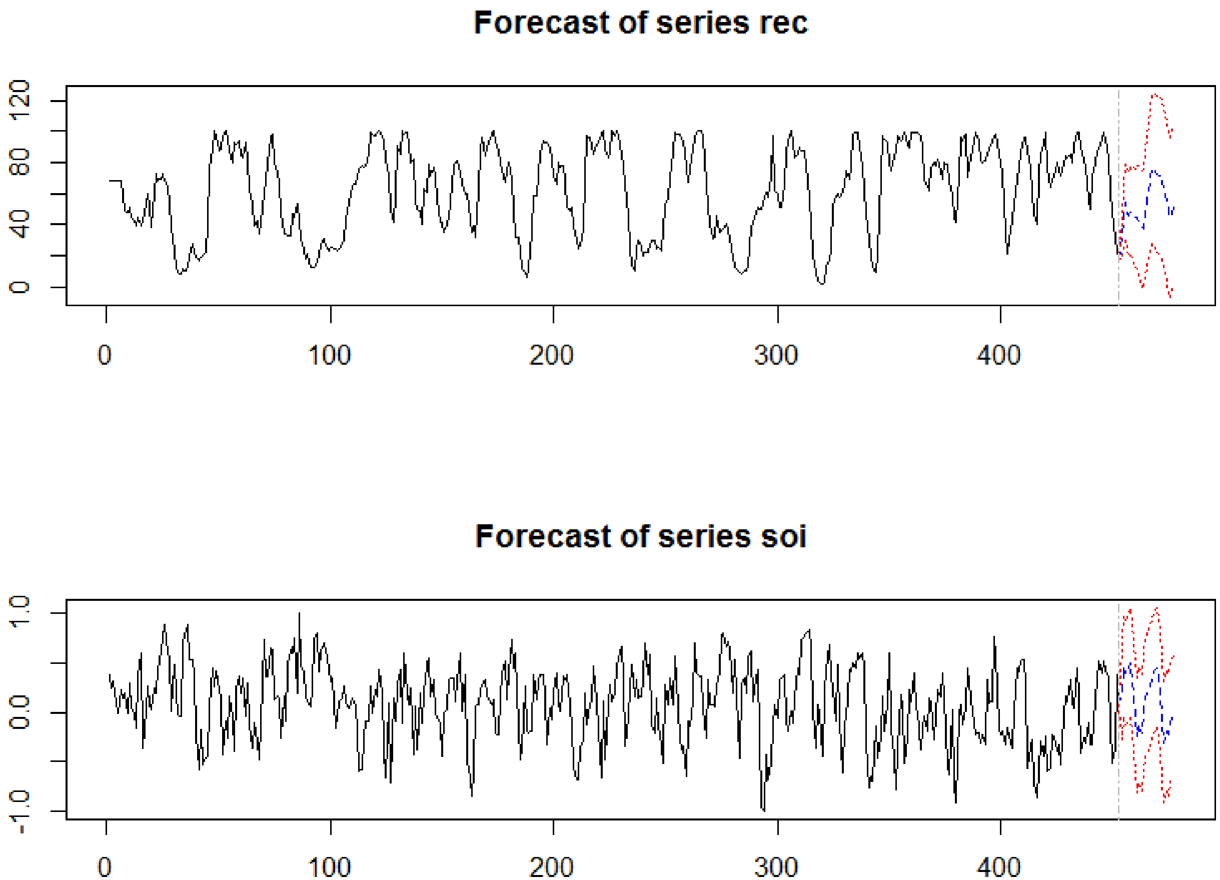
Example (cont'd)

- VAR(15) estimation

```
> VAR(X,15)

VAR Estimation Results:
=======================

Estimated coefficients for equation rec:
========================================
Call:
rec = rec.l1 + soi.l1 + rec.l2 + soi.l2 + rec.l3 + soi.l3 + rec.l4 + soi.l4 + rec.l5 + soi.l5 + rec.l6 +
soi.l6 + rec.l7 + soi.l7 + rec.l8 + soi.l8 + rec.l9 + soi.l9 + rec.l10 + soi.l10 + rec.l11 + soi.l11 + r
ec.l12 + soi.l12 + rec.l13 + soi.l13 + rec.l14 + soi.l14 + rec.l15 + soi.l15 + const

      rec.l1        soi.l1        rec.l2        soi.l2        rec.l3        soi.l3        rec.l4
 1.207691796   1.174911739  -0.393793460   0.433512840   0.012686405  -1.393827810  -0.149158384
      soi.l4        rec.l5        soi.l5        rec.l6        soi.l6        rec.l7        soi.l7
 0.019510371   0.199015432 -21.480745005   0.017758950   9.115855299  -0.209012841  -1.888180940
      rec.l8        soi.l8        rec.l9        soi.l9       rec.l10       soi.l10       rec.l11
 0.195335535  -2.403517603  -0.121119906  -2.774691973  -0.011415651  -0.093660952  -0.003666367
     soi.l11       rec.l12       soi.l12       rec.l13       soi.l13       rec.l14       soi.l14
-0.711908788   0.060769283  -4.103027176  -0.039959051   2.923913383  -0.041878162  -1.594892567
     rec.l15       soi.l15         const
 0.001724510  -0.826725126  19.000121895
```

Example (cont'd)

- Predictions w/ `predict()` function



Granger Causality

- Consider 2D VAR(1) model

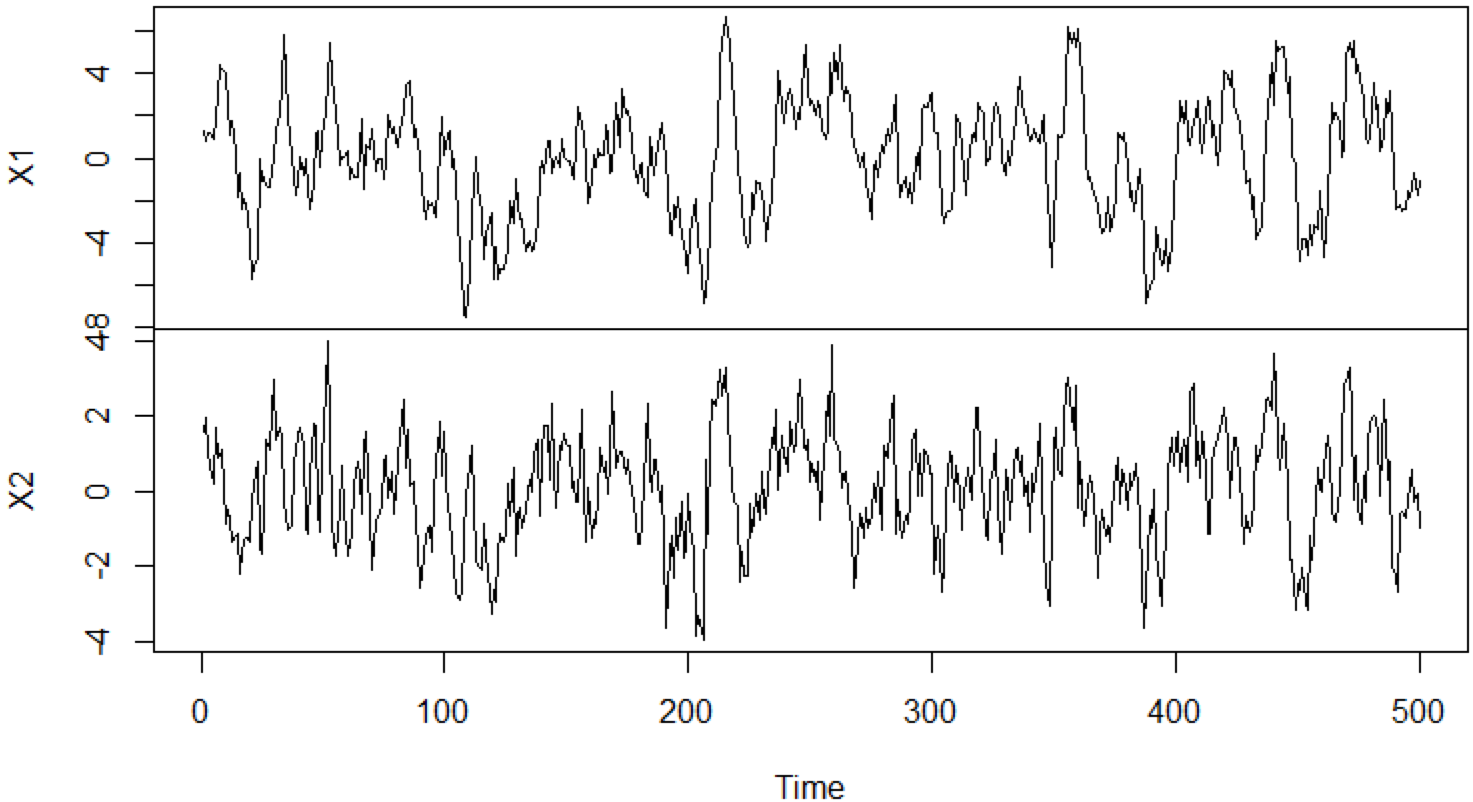
$$\begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} = \begin{bmatrix} \varphi_{1,1} & \varphi_{1,2} \\ 0 & \varphi_{2,2} \end{bmatrix} \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \begin{bmatrix} W_{1,t} \\ W_{2,t} \end{bmatrix} \Leftrightarrow$$

$$\Leftrightarrow \left\{ \begin{matrix} X_{1,t} = \varphi_{1,1} X_{1,t-1} + \varphi_{1,2} X_{1,t-2} + W_{1,t} \\ X_{2,t} = \varphi_{2,2} X_{2,t-1} + W_{2,t} \end{matrix} \right\}$$

  - Coordinate $X_{1,t}$ depends on *both* $X_{1,t-1}$ & $X_{2,t-1}$, but coordinate $X_{2,t}$ depends on $X_{2,t-1}$ *only*
  - Nevertheless, both coordinates can be cross-correlated at different lags

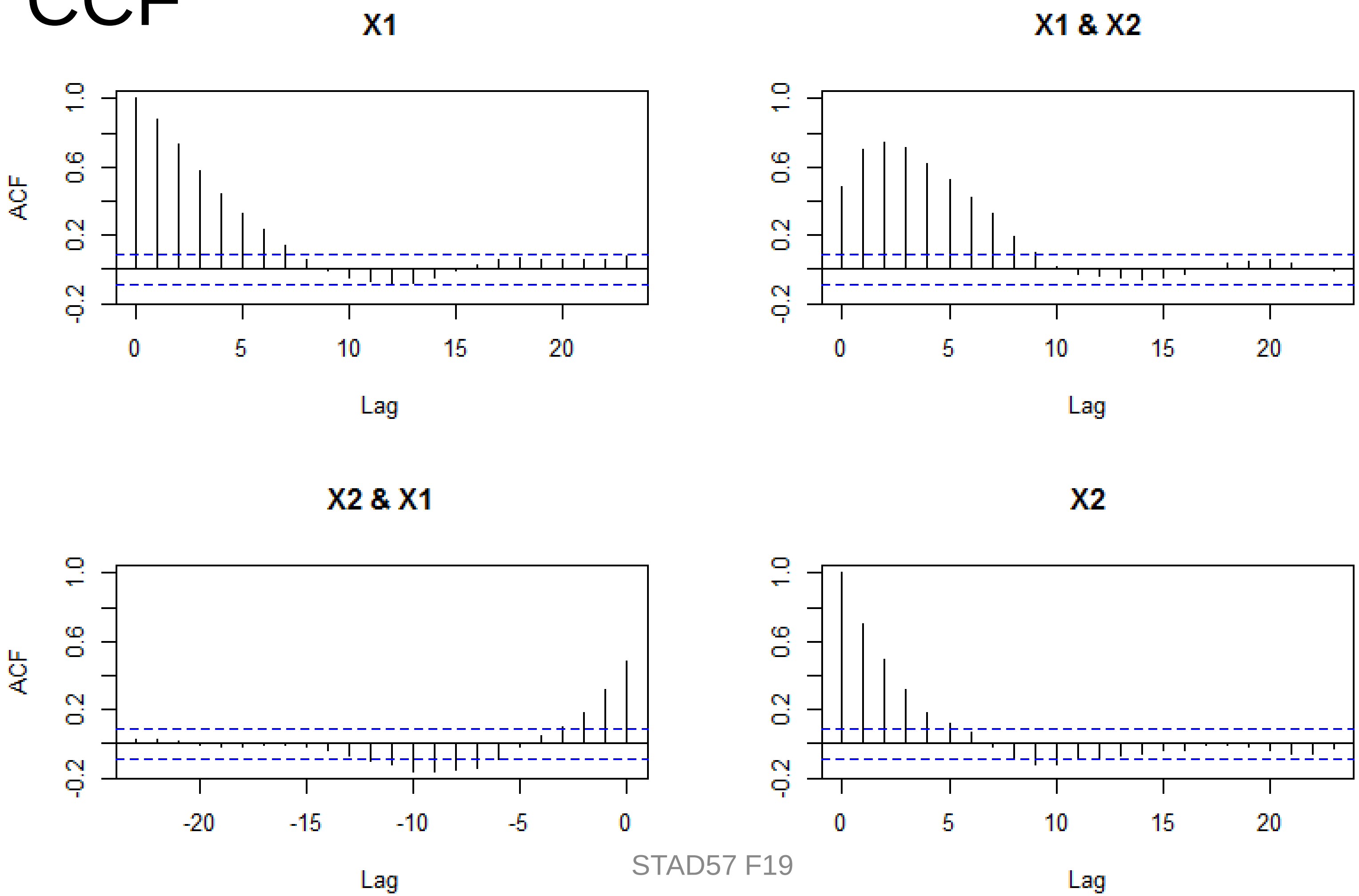
Example

- Simulated series from $\begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} = \begin{bmatrix} .7 & .7 \\ 0 & .7 \end{bmatrix} \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \begin{bmatrix} W_{1,t} \\ W_{2,t} \end{bmatrix}$



Example (cont'd)

- CCF



Granger Causality

- TS $\{Y_t\}$ is said to *Granger-cause* TS $\{X_t\}$ if past of $\{Y_t\}$ helps in predicting $\{X_t\}$ beyond using past of $\{X_t\}$ alone
  - Clive Granger, 2003 Nobel prize in Economics
- In terms of VAR(p) model, Granger-causality implies certain structure of zero-coefficients in the dynamic equation

Example

- For VAR(1) $\begin{bmatrix} X_{1,t} \\ X_{2,t} \\ X_{3,t} \end{bmatrix} = \begin{bmatrix} \varphi_{1;1,1} & \varphi_{1;1,2} & 0 \\ 0 & \varphi_{1;2,2} & \varphi_{1;2,3} \\ 0 & 0 & \varphi_{1;3,3} \end{bmatrix} \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \\ X_{3,t-1} \end{bmatrix} + \begin{bmatrix} W_{1,t} \\ W_{2,t} \\ W_{3,t} \end{bmatrix}$

  find which variable Granger-causes which

For predicting $\{X_1\}$: $\{X_2\}$ Granger causes $\{X_1\}$ (b/c of $\varphi_{1,2}$)
$\{X_3\}$ does not Granger-cause $\{X_1\}$ (b/c given (X_1, X_2), then X_3 is not used for prediction)

For predicting $\{X_2\}$: $\{X_1\}$ does not Granger cause $\{X_2\}$
$\{X_3\}$ Granger causes $\{X_2\}$

For predictin $\{X_3\}$: neither $\{X_1\}$ nor $\{X_2\}$ Granger causes $\{X_3\}$

Granger Causality

- Granger-causality based on VAR(p) model

$$\underbrace{\texttt{causality(VAR.fit}}_{\text{VAR() output}}\texttt{, }\underbrace{\texttt{cause="X1"}}_{\text{cause TS name}}\texttt{)}$$

  - R output:

```
> causality(out, cause='X1')
$Granger

 Granger causality H0: X1 do not Granger-cause X2

data:  VAR object out
F-Test = 0.52434, df1 = 1, df2 = 992, p-value = 0.4692


$Instant

 H0: No instantaneous causality between: X1 and X2

data:  VAR object out
Chi-squared = 0.27869, df = 1, p-value = 0.5976
```

```
> causality(out, cause='X2')
$Granger

 Granger causality H0: X2 do not Granger-cause X1

data:  VAR object out
F-Test = 366.14, df1 = 1, df2 = 992, p-value < 2.2e-16


$Instant

 H0: No instantaneous causality between: X2 and X1

data:  VAR object out
Chi-squared = 0.27869, df = 1, p-value = 0.5976
```

Impulse Response Function

- Want to know how changes in one coordinate affect others
  - Assume: $$\begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \begin{bmatrix} W_{1,t} \\ W_{2,t} \end{bmatrix}$$
  - Let $$\begin{bmatrix} X_{1,0} \\ X_{2,0} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \;\&\; \begin{bmatrix} W_{1,1} \\ W_{2,1} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$
  - How does 1unit change in $W_{1,\mathrm{t}}$ propagate through time?

Example

$$\begin{bmatrix} x_{1,1} \\ X_{2,1} \end{bmatrix} = \begin{bmatrix} 1/2 & 1/2 \\ 0 & 1/2 \end{bmatrix} \cdot \begin{bmatrix} X_{1,0} = 0 \\ X_{2,0} = 0 \end{bmatrix} + \begin{bmatrix} w_{1,1} = 1 \\ w_{2,1} = 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} X_{1,2} \\ X_{2,2} \end{bmatrix} = \begin{bmatrix} 1/2 & 1/2 \\ 0 & 1/2 \end{bmatrix} \cdot \begin{bmatrix} X_{1,1} = 1 \\ X_{2,1} = 0 \end{bmatrix} + \begin{bmatrix} W_{1,2} = 0 \\ W_{2,2} = 0 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} X_{1,3} \\ X_{2,3} \end{bmatrix} = \begin{bmatrix} 1/2 & 1/2 \\ 0 & 1/2 \end{bmatrix} \cdot \underbrace{\begin{bmatrix} X_{1,2} \\ X_{2,2} \end{bmatrix}} + \underline{0} = \begin{bmatrix} 1/2 & 1/2 \\ 0 & 1/2 \end{bmatrix} \cdot \begin{bmatrix} 1/2 & 1/2 \\ 0 & 1/2 \end{bmatrix} \cdot \begin{bmatrix} X_{1,1} = 1 \\ X_{2,1} = 0 \end{bmatrix}$$

$$\vdots$$

$$\begin{bmatrix} X_{1,t+1} \\ X_{2,t+1} \end{bmatrix} = \begin{bmatrix} 1/2 & 1/2 \\ 0 & 1/2 \end{bmatrix}^t \cdot \begin{bmatrix} x_{t,1} = 1 \\ X_{2,1} = 0 \end{bmatrix} \Rightarrow \left(\text{for VAR(1):} \quad \underline{X}_{t+1} = \underline{\Phi}^t \cdot \underline{X}_1 \right)$$

Impulse Response Function

- Use causal (Wold) representation of VAR(p) model to trace effect of impulse

$$\mathbf{X}_t = \mathbf{\Phi}_1 \mathbf{X}_{t-1} + \cdots + \mathbf{\Phi}_p \mathbf{X}_{t-p} \Leftrightarrow$$

$$\mathbf{X}_t = \mathbf{W}_t + \mathbf{\Psi}_1 \mathbf{W}_{t-1} + \mathbf{\Psi}_2 \mathbf{W}_{t-2} + \cdots = \sum_{j=0}^{\infty} \mathbf{\Psi}_j \mathbf{W}_{t-j}$$

where $\mathbf{\Psi}_k = \sum_{j=0}^{\min(k,p)} \mathbf{\Psi}_{k-j} \mathbf{\Phi}_j$ & $\mathbf{\Psi}_0 = \mathbf{I}$

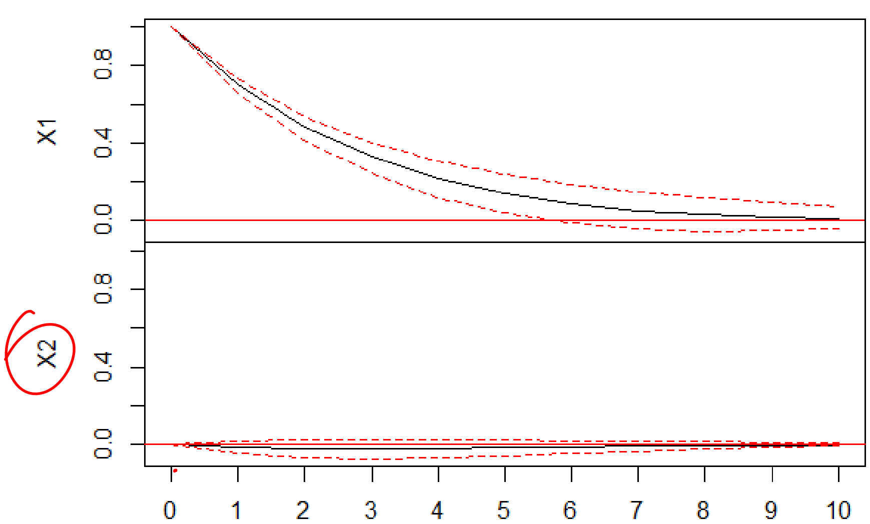
- Impulse Response Function (IRF) is given by components of $\mathbf{\Psi}$-matrices

Example

$w_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

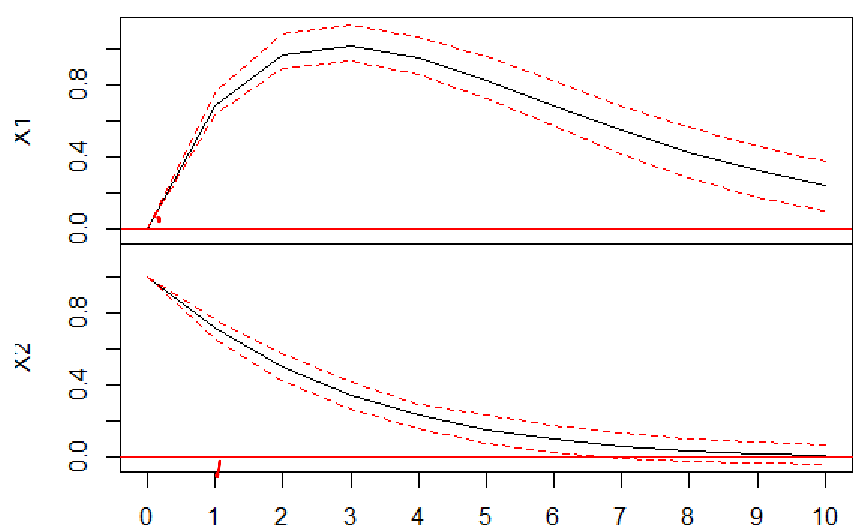
$w_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

Impulse Response from X1



95 % Bootstrap CI, 100 runs

Impulse Response from X2



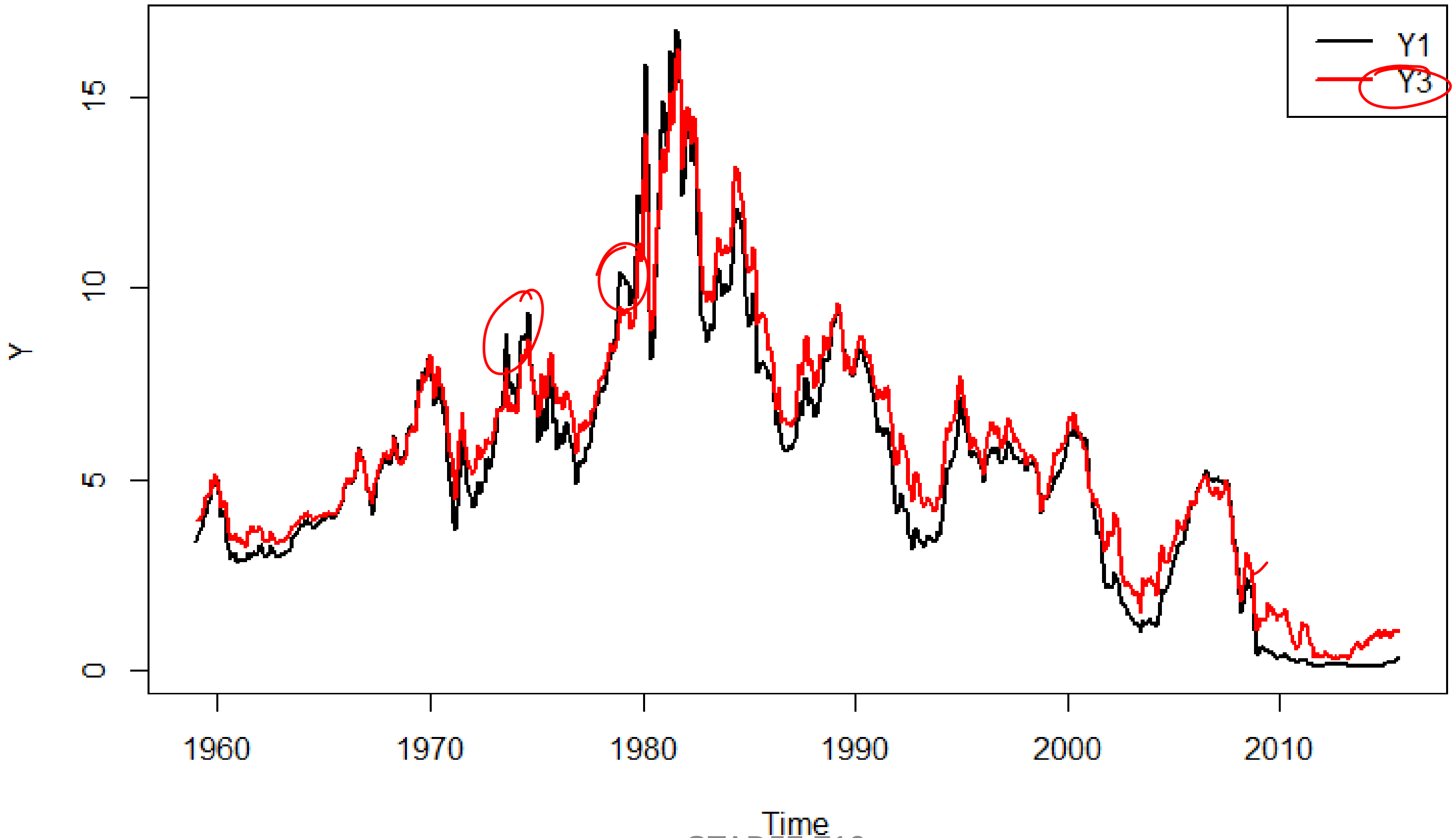
95 % Bootstrap CI, 100 runs

Cointegration

- Set of TS called *cointegrated* if :
  - Individual TS are *integrated*
    - e.g. follow $I(1)$~random walk
  - Some *linear combination* thereof is *stationary*
- E.g. Term-structure of interest rates
  - Consider yield rates of Gov't issued bonds with different maturities: e.g. 1yr vs 3yr
  - Interest levels fluctuate like a random walk, but rates for different maturities are close

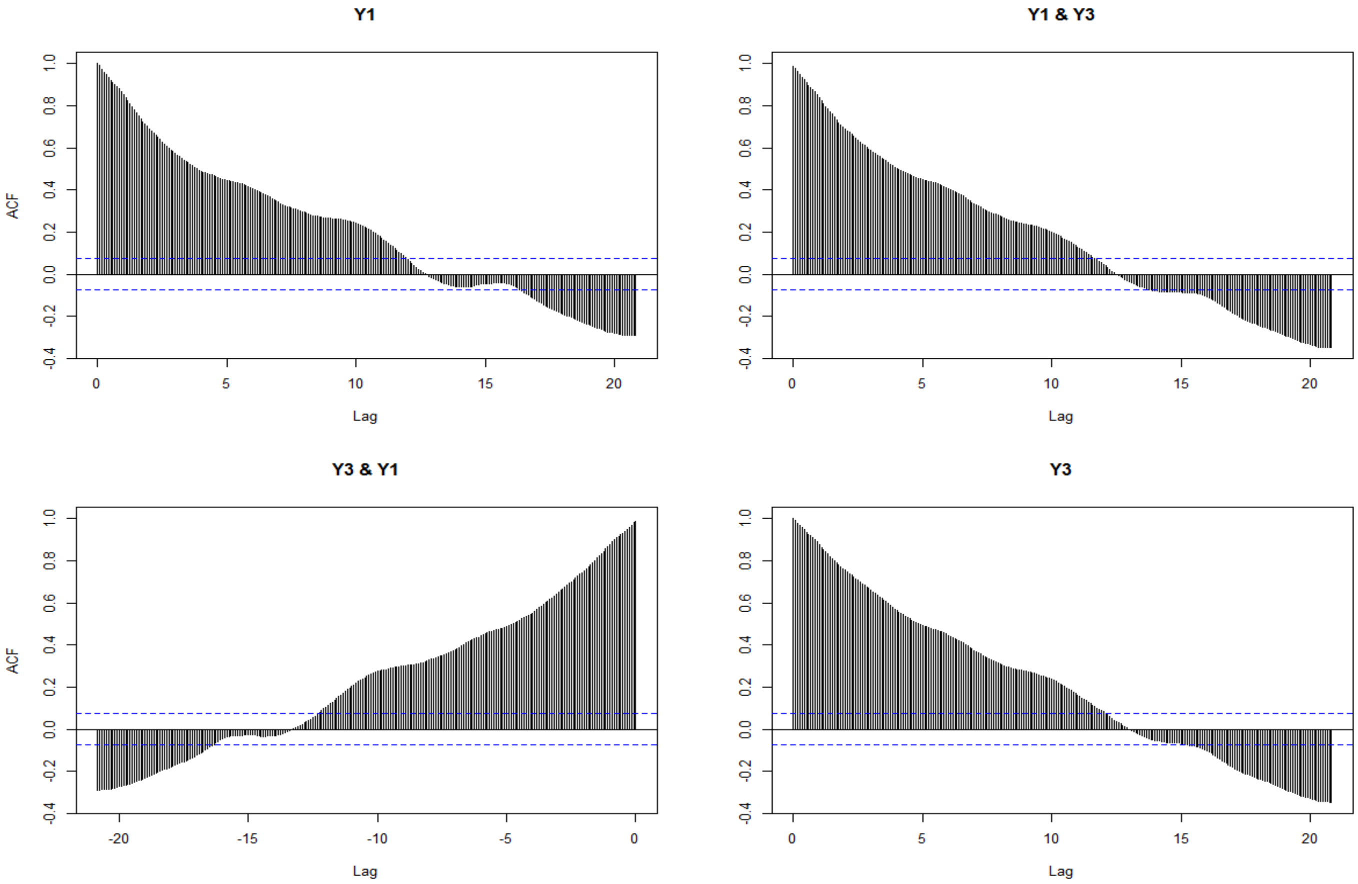
Example

- 1- & 3-year US Gov't bond yield rates



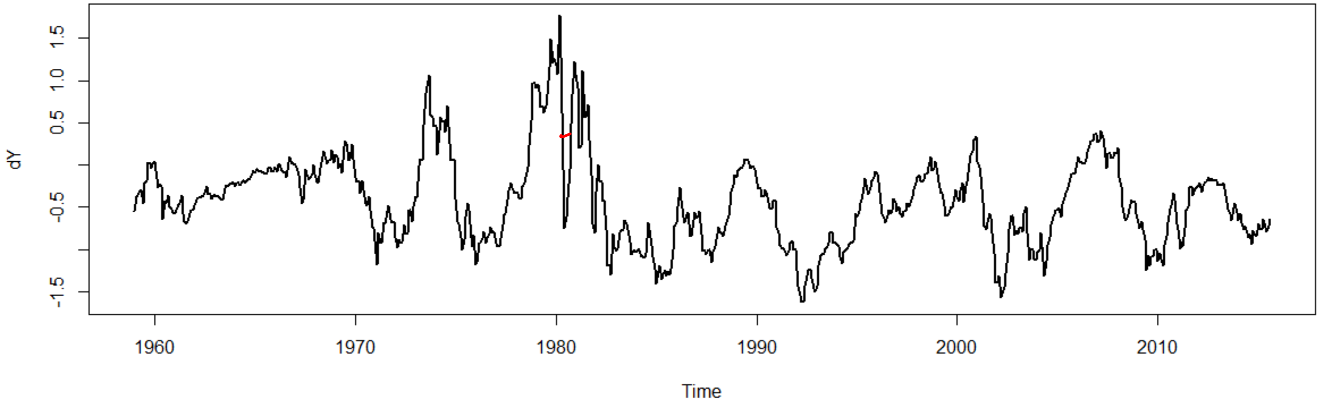
Example (cont'd)

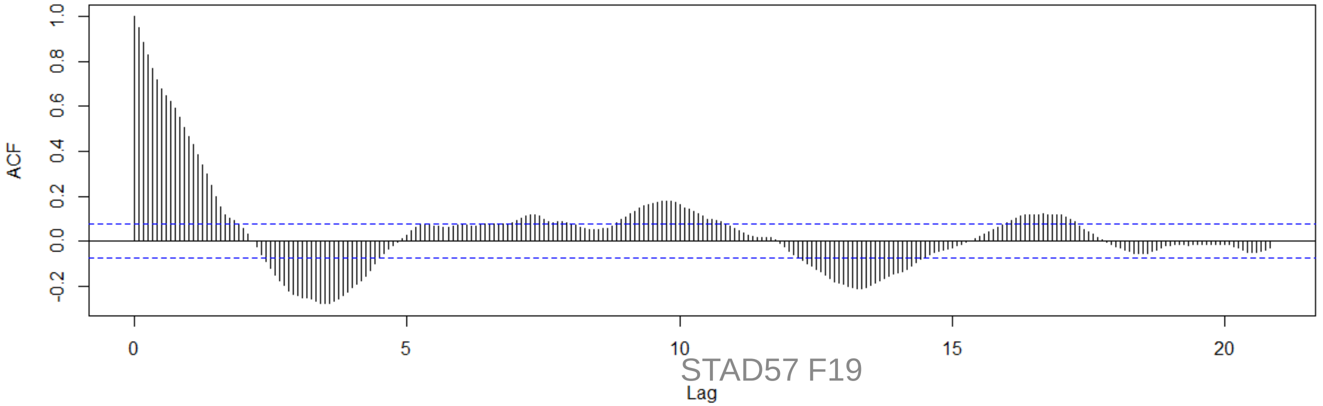
- ACF/CCF



Example (cont'd)

- Difference Y1-Y3





Cointegration

- If you *know stationary relation*, just test it for stationarity (w/ unit root tests)
  - E.g. Augmented Dickey Fuller (ADF) test
- If you *don't know stationary relation*, need to estimate; two approaches
  - Engle-Granger two-step process
    1. Regression to estimate stationary relation
    2. Perform unit root test on residuals
  - Johansen test, using VAR models

Example (cont'd)

- Unit root tests (ADF) on Y1, Y3 & Y1-Y3

```
> adf.test(Y[,"Y1"])

 Augmented Dickey-Fuller Test

data:  Y[, "Y1"]
Dickey-Fuller = -2.0811, Lag order = 8, p-value = 0.544
alternative hypothesis: stationary
```

$H_0: \varphi_1 = 1$

$H_1: \varphi_1 < 1$

; don't reject H_0 ⇒ integrated

```
> adf.test(Y[,"Y3"])

 Augmented Dickey-Fuller Test

data:  Y[, "Y3"]
Dickey-Fuller = -1.8658, Lag order = 8, p-value = 0.6351
alternative hypothesis: stationary

> adf.test(dY)

 Augmented Dickey-Fuller Test

data:  dY
Dickey-Fuller = -3.9589, Lag order = 8, p-value = 0.01105
alternative hypothesis: stationary
```

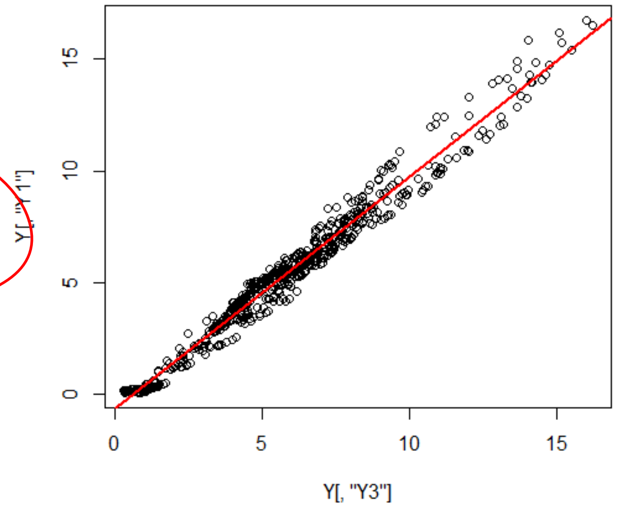
$\leq 5\%$

Example (cont'd)

- Engle-Granger

```
> (out=lm(Y1~Y3,data=Y))

Call:
lm(formula = Y1 ~ Y3, data = Y)

Coefficients:
(Intercept)           Y3
    -0.6348       1.0363

> dY.est=residuals(out)
> adf.test(dY.est)

 Augmented Dickey-Fuller Test

data:  dY.est
Dickey-Fuller = -3.9004, Lag order = 8, p-value = 0.01398
alternative hypothesis: stationary
```

$Y_1 - 1.0363\,Y_3 + .6348$



Spurious Regression

- Consider *independent* random walks $\{W_t, V_t\}$
  - When you regress $W_t = \beta_0 + \beta V_t + e_t, \ t = 1, \ldots, n$ you are NOT guaranteed that $\hat{\beta} \to 0$ as the sample size $n \to \infty$ (i.e. not consistent)!!!

- Effect called *spurious* (fake) regression
  - Results of random walk (integrated series) regressions are NOT reliable

Cointegration & VAR

- Consider VAR (p)

$$\mathbf{X}_t = \mathbf{\Phi}_1 \mathbf{X}_{t-1} + \cdots + \mathbf{\Phi}_p \mathbf{X}_{t-p} + \mathbf{W}_t$$

  - Model is stable (causal) if

$$\det\left(\mathbf{I} - \mathbf{\Phi}_1 z - \cdots - \mathbf{\Phi}_p z^p\right) \neq 0, \;\; \forall \, |\, z \,| \leq 1$$

  - If there is a *unit root*, then all or some of the coordinates of $\mathbf{X}_t$ are $I(1)$
- If model is cointegrated, some linear combination of $\mathbf{X}_t$ are $I(0)$

Cointegration & VAR

- Write VAR model as *Vector Error Correction Model* (*VECM*)

$$\Delta \mathbf{X}_t = \mathbf{\Lambda X}_{t-1} + \mathbf{\Lambda}_1 \Delta \mathbf{X}_{t-1} + \cdots + \mathbf{\Lambda}_{p-1} \Delta \mathbf{X}_{t-p+1} + \mathbf{W}_t \Rightarrow$$

- where $\begin{cases} \mathbf{\Lambda} = \mathbf{\Phi}_1 + \cdots + \mathbf{\Phi}_p - \mathbf{I} \\ \mathbf{\Lambda}_i = -(\mathbf{\Phi}_{i+1} + \cdots + \mathbf{\Phi}_p) = \sum_{k=i+1}^{p} \mathbf{\Phi}_k \end{cases}$

$X_t \sim I(1)$. Show VECM for $\Delta X_t \iff$ VAR for X_t

has unit root

$$\Delta X_t = \underbrace{\Lambda X_{t-1}}_{} + \Lambda_1 \Delta X_{t-1} + \dots + \Lambda_{p-1} \Delta X_{t-p+1} + W_t$$

$$= (\Phi_1 + \cancel{\Phi_2} + \cancel{\Phi_3} + \dots + \cancel{\Phi_p} - I) \cdot X_{t-1}$$

$$- (\Phi_2 + \cancel{\Phi_3} + \dots + \cancel{\Phi_p}) \cdot (\cancel{X_{t-1}} - X_{t-2})$$

$$- (\Phi_3 + \dots + \cancel{\Phi_p}) \cdot (\cancel{X_{t-2}} - X_{t-3})$$

$$\vdots$$

$$- \Phi_p \cdot (\cancel{X_{t-p+1}} - X_{t-p}) + W_t$$

$$\Rightarrow X_t - \cancel{X_{t-1}} = \Phi_1 X_{t-1} - \cancel{X_{t-1}} + \Phi_2 X_{t-2} + \dots + \Phi_p X_{t-p} + W_t$$

$$X_t = \sum_{j=1}^{p} \Phi_j X_{t-j} + W_t \;) \Rightarrow \text{VAR}(p)$$

Cointegration & VAR

- For the VECM

$$\Delta\mathbf{X}_t = \mathbf{\Lambda}\mathbf{X}_{t-1} + \mathbf{\Lambda}_1\Delta\mathbf{X}_{t-1} + \cdots + \mathbf{\Lambda}_{p-1}\Delta\mathbf{X}_{t-p+1} + \mathbf{W}_t, \quad \mathbf{X} \in \mathbb{R}^d$$

  - $\{\Delta\mathbf{X}_t\}$ is $I(0)$, but $\{\mathbf{X}_t\}$ is $I(1)$
  - Thus, term $\mathbf{\Lambda}\mathbf{X}_{t-1}$ must also be $I(0) \rightarrow \mathbf{\Lambda}$ must contain cointegration relation(s)
  - Since $\det\left(\mathbf{I} - \mathbf{\Phi}_1 - \cdots - \mathbf{\Phi}_p\right) = \det\left(\mathbf{\Lambda}\right) = 0$ from unit root of $\{\mathbf{X}\}$, matrix $\mathbf{\Lambda}$ has *reduced rank* ($r<d$), i.e can be written as $\underset{(d\times d)}{\mathbf{\Lambda}} = \underset{(d\times r)}{\boldsymbol{\alpha}} \underset{(r\times d)}{\boldsymbol{\beta}^\top}$, where β defines cointegrating relations

Example

- For $\begin{cases} \nabla X_{1,t} = \varphi_1\left(X_{1,t-1} - \lambda X_{2,t-1}\right) + \varepsilon_{1,t} \\ \nabla X_{2,t} = \varphi_2\left(X_{1,t-1} - \lambda X_{2,t-1}\right) + \varepsilon_{2,t} \end{cases}$, show that

  $Y_t = X_{1,t} - \lambda X_{2,t}$ follows AR(1) process

$$\nabla \underline{X}_t = \begin{bmatrix}\varphi_1 \\ \varphi_2\end{bmatrix} [1 \;\; -\lambda] \, \underline{X}_{t-1}$$

Take 1st line & subtract $\lambda \times$ 2nd line:

$$\nabla X_{1,t} - \lambda \nabla X_{2,t} = \varphi_1(X_{1,t-1} - \lambda X_{2,t-1}) + \varepsilon_1 - \lambda\varphi_2(X_{1,t-1} - \lambda X_{2,t-1})$$

$$\Rightarrow (\varphi_1 - \lambda\varphi_2) \cdot X_{1,t-1} - \lambda(\varphi_1 - \lambda\varphi_2) X_{2,t-1} + \varepsilon_{1,t} - \lambda\varepsilon_{2,t}$$

$$\Rightarrow (\varphi_1 - \lambda\varphi_2) \cdot (X_{1,t-1} - \lambda X_{2,t-1})$$

$$\nabla X_{1,t} - \lambda \nabla X_{2,t} = (\varphi_1 - \lambda \varphi_2)(X_{1,t-1} - \lambda X_{2,t-1}) + \underbrace{\varepsilon_{1,t} - \lambda \varepsilon_{2,t}}_{\varepsilon_t}$$

$$\Rightarrow \underbrace{(X_{1,t} - \lambda X_{2,t})}_{Y_t} - (\underbrace{X_{1,t-1} - \lambda X_{2,t-1}}_{Y_{t-1}}) =$$

$$= (\varphi_1 - \lambda \varphi_2)(\underbrace{X_{1,t-1} - \lambda X_{2,t-1}}_{Y_{t-1}}) + \varepsilon_t$$

$$\Rightarrow Y_t = \underbrace{(1 + \varphi_1 - \lambda \varphi_2)}_{= \varphi} Y_{t-1} + \varepsilon_t$$

$Y_t = \varphi Y_{t-1} + \varepsilon_t$, which is stationary $|\varphi| < 1$

Cointegration & VAR

- Johansen procedure:
  - Specify and estimate VAR(p) model for $\{\mathbf{X}_t\}$
  - Construct Likelihood Ratio(LR) tests for the rank of $\mathbf{\Lambda}$, to determine number of cointegrating vectors
    - If necessary, impose normalization and identifying restrictions on the cointegrating vectors.
  - Given cointegrating vectors, estimate resulting VECM by maximum likelihood

Example

- Find # of cointegrating vectors

```
> out=ca.jo(Y, ecdet="const", K=3)
> summary(out)

######################
# Johansen-Procedure #
######################

Test type: maximal eigenvalue statistic (lambda max) , without linear trend and constant in cointegratio
n

Eigenvalues (lambda):
[1] 3.930817e-02 3.907291e-03 4.336809e-18

Values of teststatistic and critical values of test:

          test 10pct  5pct  1pct
r <= 1 |  2.65  7.52  9.24 12.97
r = 0  | 27.15 13.75 15.67 20.20

Eigenvectors, normalised to first column:
(These are the cointegration relations)

              Y1.l3     Y3.l3  constant
Y1.l3     1.0000000  1.000000  1.000000
Y3.l3    -1.0159606 -2.429144 -1.729806
constant  0.5080471  7.914550 18.018859
```

Example (cont'd)

- Fit VECM model

$$\Lambda = \alpha \cdot \beta^T = \begin{bmatrix} -.0244 \\ .03177 \end{bmatrix} \begin{bmatrix} 1 & -1.0159 \end{bmatrix}$$

```
> cajorls(out,r=1)
$rlm

Call:
lm(formula = substitute(form1), data = data.mat)

Coefficients:
         Y1.d      Y3.d
ect1     -0.02448   0.03177
Y1.dl1    0.20721   0.05366
Y3.dl1    0.30138   0.35856
Y1.dl2   -0.16093  -0.02903
Y3.dl2   -0.09866  -0.20990


$beta
              ect1
Y1.l3     1.0000000
Y3.l3    -1.0159606
constant  0.5080471
```

α

Λ_1

Λ_2

β

$$\Delta X_t = \mu + \Lambda X_{t-1} + \Lambda_1 \Delta X_{t-1} + \Lambda_2 \Delta X_{t-2}$$