ENV797 - TIME SERIES ANALYSIS FOR ENERGY AND ENVIRONMENT APPLICATIONS

M6 - Seasonal ARIMA Models

Prof. Luana Medeiros Marangon Lima, Ph.D.

Nicholas School of the Environment - Duke University

Learning goals

- Discuss the seasonal ARIMA model – SARIMA

- Learn how to identify order of seasonal component

- Learn how to fit SARIMA models in R

Seasonal Models

SARIMA

Modeling with Seasonality

- So far we have two ways to deal with seasonality
  - Linear regression on seasonal dummy variables and work with residuals
  - Computing and subtracting the average of each "month" (season)

- In R: use stats::decompose() and forecast::seasadj()

- Drawbacks
  - Need to add seasonality back after fitting the model
  - Seasonal component is modeled as "constant" over time

Seasonal Models

□ Today we will learn a new approach: work with seasonal models, i.e., models that can handle seasonal time series

□ We still need to identify the seasonal component but we do not need to remove it from the series
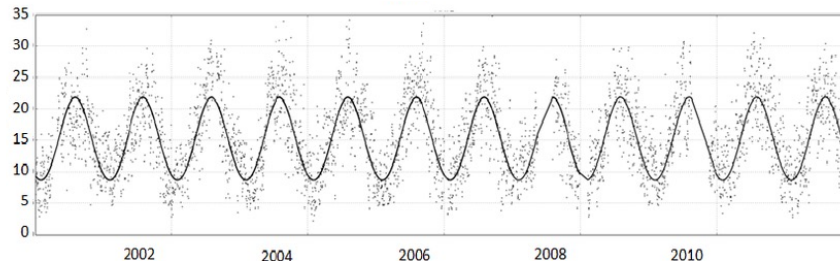
□ And we don't need to add it back after fitting the model

Seasonal ARIMA Models

- The Seasonal ARIMA models rely on seasonal lags and difference to fit the seasonal pattern

- The seasonal part of the ARIMA model have three additional parameters

> P: *number of seasonal autoregressive terms*
> D: *number of seasonal differences*
> Q: *number of moving average terms*

- The complete model is called an

$$\mathbf{ARIMA}(p, d, q) \times (P, D, Q)_s$$

Non-seasonal part   Seasonal part

Seasonal ARIMA Models (cont'd)

- We have two terms that refer to differencing

- How does all this differencing work?

$$If\ d = 0, D = 1: \qquad y_t = Y_t - Y_{t-s}$$
$$If\ d = 1, D = 1: \qquad y_t = (Y_t - Y_{t-1}) - (Y_{t-s} - Y_{t-s-1})$$

- Note that s is the seasonal period, i.e., for monthly data $s = 12$

- D should never be more than 1

- $(d + D)$ should never be more than 2

Seasonal ARIMA Models (cont'd)

- What about the seasonal AR and MA terms?

  $If\ P = 1:$ y_{t-s} is added to the equation P always refer to autoregressive

  $If\ Q = 1:$ a_{t-s} is added to the equation Q is the moving average.

- $(P + Q)$ should not be more than one

- **After differencing** plot ACF and PACF

  - Positive spikes in ACF at lag 12, 24, 36, …
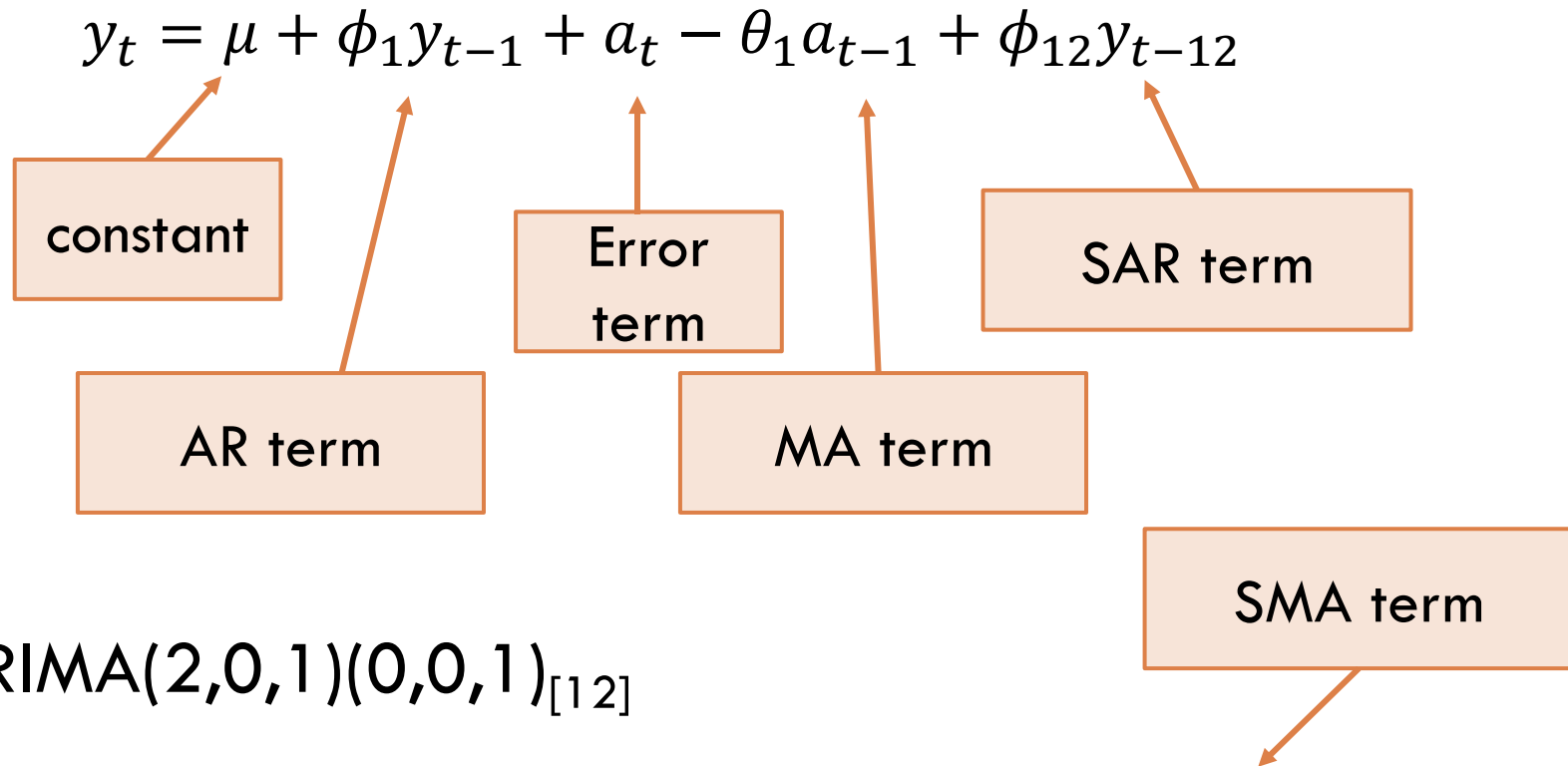
    and single positive spike in PACF at lag 12 $P = 1$

  - Negative spike in ACF at lag 12

    and negative spikes in PACF at lag 12, 24, 36, …   Q= 1

Examples: Seasonal ARIMA

□ SARIMA$(1,0,1)(1,0,0)_{[12]}$ If you difference once, you can have a drift, or no drift otherwise.

$$y_t = \mu + \phi_1 y_{t-1} + a_t - \theta_1 a_{t-1} + \phi_{12} y_{t-12}$$

constant

Error term

SAR term

AR term

MA term

SMA term

□ SARIMA$(2,0,1)(0,0,1)_{[12]}$

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + a_t - \theta_1 a_{t-1} - \theta_{12} a_{t-12}$$

Periodic AR Model (PAR)

Fits a model for each season of the year for any order of choice.

☐ Also known as PAR(p) where p is the order of autoregressive model

☐ Similar to having one AR model for each season of the year

$$y_t = \phi_{1,s} y_{t-1} + \cdots + \phi_{p,s} y_{t-p} + a_t$$

☐ Note that $s = 1,2,3,\dots$ depending on the number of seasons

☐ The autoregressive parameter vary with the season for lag

ARIMA Models Summary

ARIMA Models Summary

- AR(p), MA(q), ARMA(p,q) – stationary and non-seasonal process

- ARIMA(p,d,q) – non-stationary and non-seasonal process

  - The Integrated part can handle the non-stationarity

  - The non-stationary could be either a deterministic or stochastic trend eliminated by differencing

- Seasonal ARIMA or SARIMA – non-stationary and seasonal process

No s for this

$$\mathbf{ARIMA}(p, d, q) \times (P, D, Q)_S$$

Non-seasonal part   Seasonal part

Fitting ARIMA Models

Cheat sheet

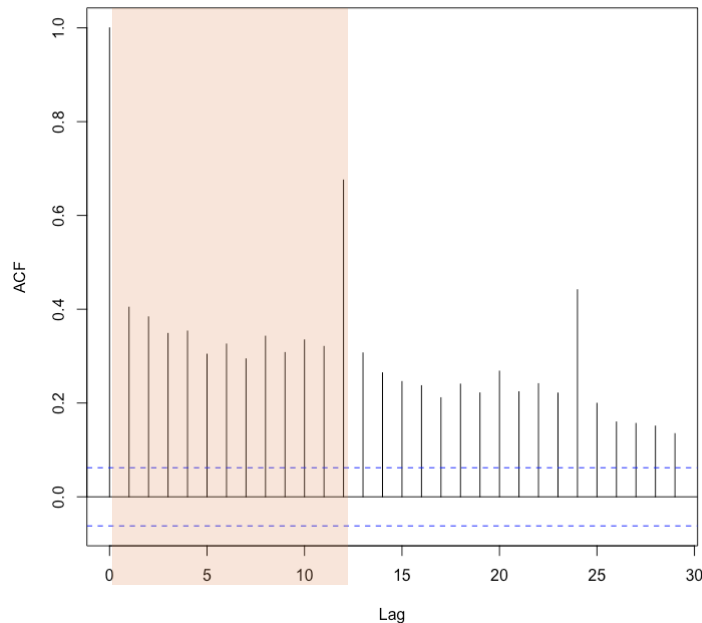# How to Specify Model Orders?

Start with non-seasonal part

□ Step 1: Run stationary test

- To identify stochastic trend use ADF
- To identify determinist trend use Mann Kendall

□ Step 2: If either trend is present $d = 1$

- It may be the case that you need differencing more than once to remove trend
- To find out if $d = 1$ is enough, run the tests again on the differenced series
- Repeat the process until there is no trend on data ($d \leq 2$)
- There is a function in R that returns the number of differences needed to achieve stationary

```
ndiffs(x, alpha = 0.05, test = c("kpss", "adf", "pp"), max.d = 2)
```
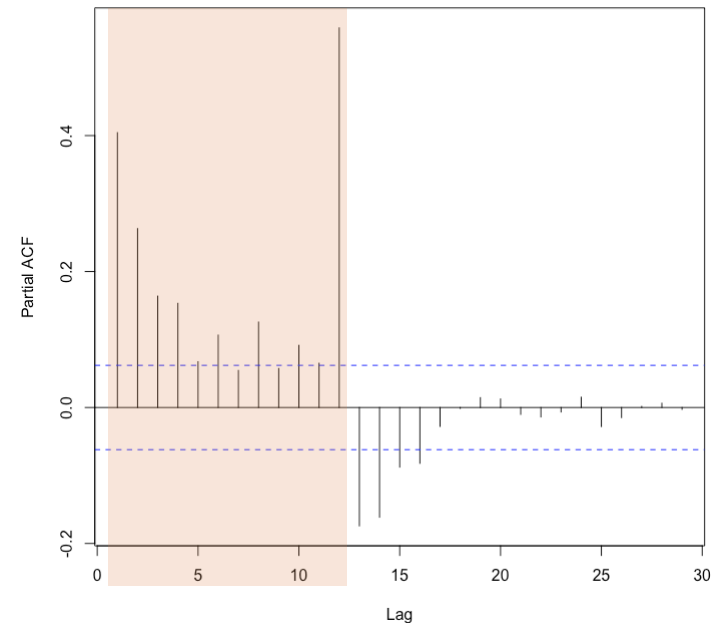x is your timeseries

How to Specify Model Orders?

☐ Step 3: Plot ACF and PACF and look the behavior of non-seasonal lags

If cuts off (MA process)
If tails off, i.e., slow decay (AR process)

ACF gives you the value of q

If cuts off (AR process)
If tails off, i.e., slow decay (MA process)

PACF gives you the value of p

# How to Specify Model Orders?

<u>Move to seasonal component</u>

☐ Step 4: Check if seasonal differencing is needed

- Seasonality can usually be verified by plotting your time series or by spikes at equally spaced lags (multiples of lag s) on ACF and PACF

- If you are still not sure about the existence of seasonal component you can run statistical tests

- The seasonal difference is needed if the seasonal pattern is strong and stable over time

- There is also a function in R that returns the number of seasonal differences needed to achieve stationary

```
nsdiffs(x, m = frequency(x), test = c("ocsb", "ch"), max.D = 1)
```
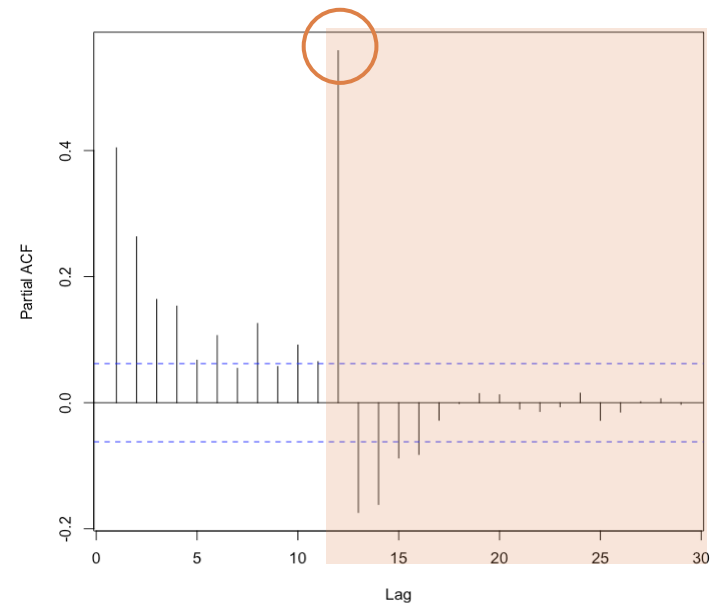
How to Specify Model Orders?

Move to seasonal component

☐ Step 5: Again ACF and PACF plots but now look at the seasonal lags
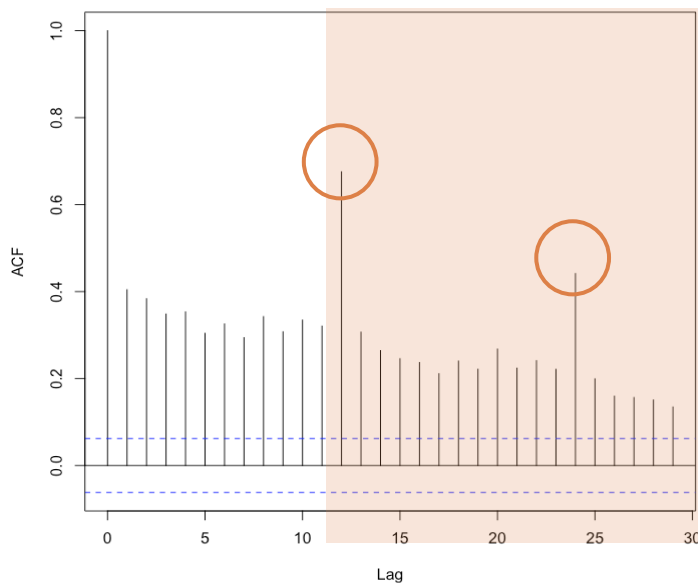
Multiple spikes at seas. lag (SAR process)
Single Spike (SMA process)

ACF gives you the value of Q

Multiple spikes at seas. lag (SMA process)
Single Spike (SAR process)

PACF gives you the value of P

Some Rules for ARIMA Modeling

- If the series have positive autocorrelations out to a high number of lags, then it probably needs differencing more than once

- If the lag-1 autocorrelation is zero or negative, or the autocorrelations are all small with no pattern, then the series does not need a higher order of differencing

- If the lag-1 autocorrelation is -0.5 or more negative, the series may be overdifferenced

Some Rules for ARIMA Modeling

- Never use more than 1 order of seasonal differencing

$$D \leq 1$$

- Never use more than 2 orders of total differencing

$$seasonal\ (D) + nonseasonal\ (d) \leq 2$$

- If the autocorrelation at the seasonal period is <u>positive</u>, consider adding an **SAR** term to the model

- If the autocorrelation at the seasonal period is <u>negative</u>, consider adding an **SMA**

- Avoid mixing SAR and SMA terms in the same model, and avoid using more than one of either kind

$$P + Q \leq 1$$

P or Q should be always equal to 1.

The ndiffs() and nsdiffs() tests

Augmented Dickey Fuller (ADF)

Check for stochastic trend

$$\begin{cases} H_0: & \phi = 1 \ (i.e. \ contain \ a \ unit \ root) \\ H_1: & \phi < 1 \ (i.e. \ is \ stationary) \end{cases}$$

Kitawoski-Phillips-Schmidt-Shin (KPSS)

Check for deterministic & stochastic trend

$$\begin{cases} H_0: & stationary \ around \ determinist \ trend \\ H_1: & contain \ a \ unit \ root \end{cases}$$

Phillips–Perron test (PP)

Check for stochastic trend

$$\begin{cases} H_0: & \phi = 1 \ (i.e. \ contain \ a \ unit \ root) \\ H_1: & \phi < 1 \ (i.e. \ is \ stationary) \end{cases}$$

Canova and Hansen (CH)

Check for seasonal pattern

$$\begin{cases} H_0: & seasonal \ pattern \ is \ stable \\ H_1: & not \ stable. \end{cases}$$

Osborn,. Chui, Smith and Birchenhall (OCSB)

Check for seasonal unit root
$$\begin{cases} H_0: & seasonal \ unit \ root. \\ H_1: & no \ seasonal \ unit \ root \end{cases}$$

Covariates in ARIMA

ARIMAX model

More on ARIMA and SARIMA in R

Constants in the model

Constants in ARIMA models in R

- The function Arima() takes three arguments regarding constants
  - include.mean – only has effect when $d = 0$ and is true by default
  - include.drift – allows $\mu \neq 0$ when $d = 1$ and is false by default
  - include.constant – more general. If TRUE, will set include.mean=TRUE if $d = 0$ and include.drift=TRUE if $d = 1$
  - Note: For d>1 no constant is allowed   This is differencing for lagging 1.
- The function auto.arima automates the inclusion of a constant
  - If alllowdrift=FALSE is specified then the constant is only allowed when $d = 0$
- The arima() function only has the include.mean option

THANK YOU !

luana.marangon.lima@duke.edu

Master of Environmental Management Program
Nicholas School of the Environment - Duke University