Time series analysis
Day 2.

David Pupovac

What will we do today

- Up until now:
  - We discussed some basic concepts related to time series analysis
  - We approached time series from multiple regression perspective and we talked about FGLS and Newey-West standard errors
  - NONE OF THESE MODELS SHOULD INCLUDE LAG DEPENDENT VARIABLE (although Wooldrige (2000) indicates that it is valid to us Newey West standard errors with lagged dependent variable). However, if you model dynamics in principal there should be no need to correct standard errors using Newey-West procedure)
- Today instead multivariate analysis we will focus only on univariate analysis
  - We will discuss some descriptive techniques which will help us in understanding time series (but we will not do modeling just yet)?
  - We will introduce some of the most fundamental concepts of time series
  - Finally, we will introduce concepts of AR (autoregressive) and MA (moving average) processes and actually start to model time series

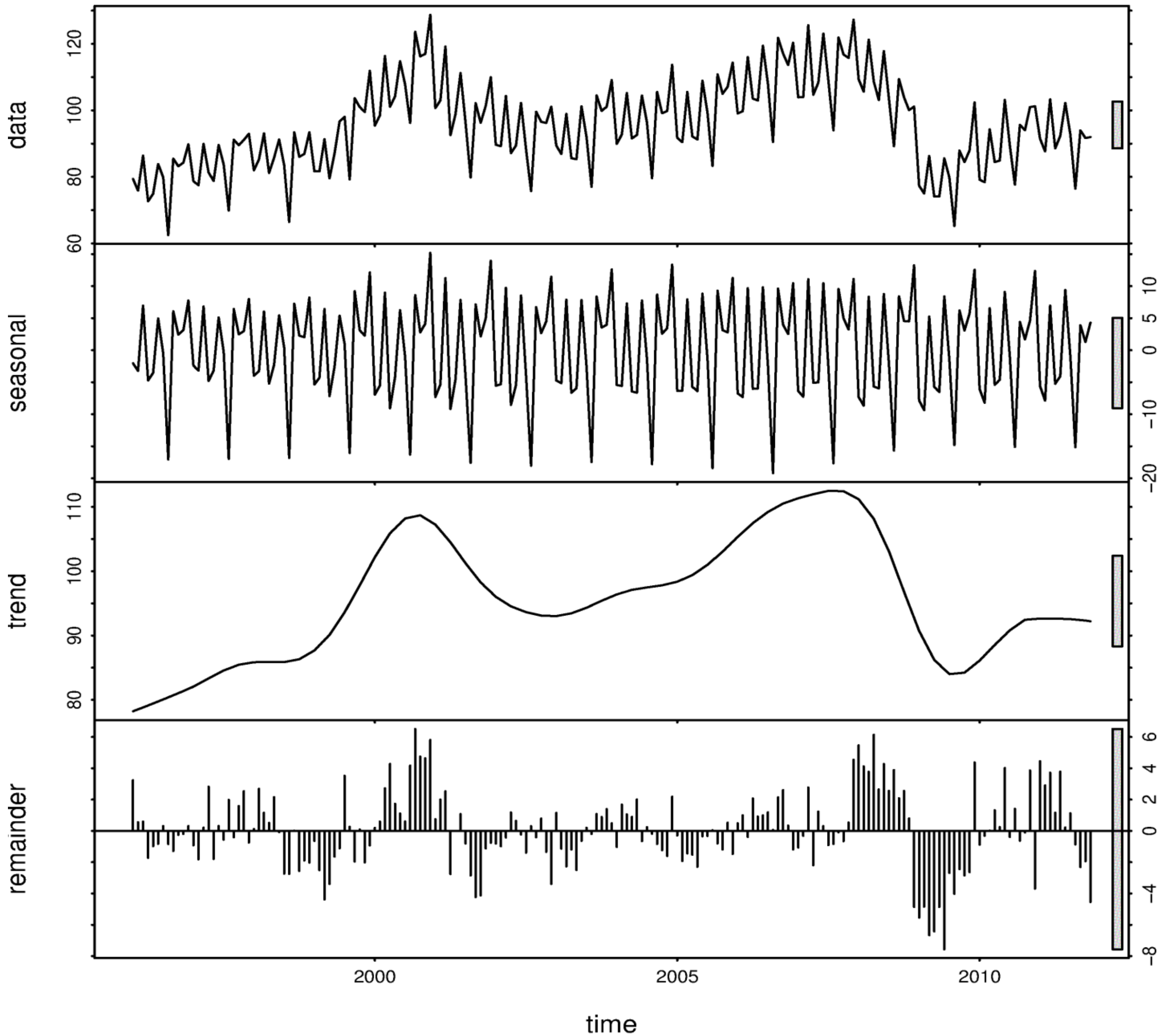
Signal

- Goal of time series is not find the straight line relationship between variable of interest and time but rather to isolate the signal - the systematic predictable component of time series from random noise

- **Three components of the signal:**

  – **Trend** - a persistent tendency for a series to increase or decrease (e.g. populations, gdp)
  – **Cycles** - oscillations around trend line or mean over several intervals, typically not periodically regular.
  – **Seasonality** - tendency of a series to increase and decrease in regular ways (daily temperature)

- **Random error**



data
seasonal
trend
remainder
time
2000
2005
2010
60
80
100
120
−20
−10
0
5
10
80
90
100
110
−8
−4
0
2
4
6

Variance decomposition

- If $\boldsymbol{y_t}$ is the observed series, $\boldsymbol{T_t}$ is the trend, $\boldsymbol{S_t}$ is the seasonal effect, and $\boldsymbol{\varepsilon_t}$ is an error term:

- **Additive model is:**

$$\boldsymbol{y_t = T_t + S_t + \varepsilon_t}$$

used when the amplitude of the seasonal effect is the same across time and the residuals are roughly the same size throughout the series

- **Multiplicative model is:**

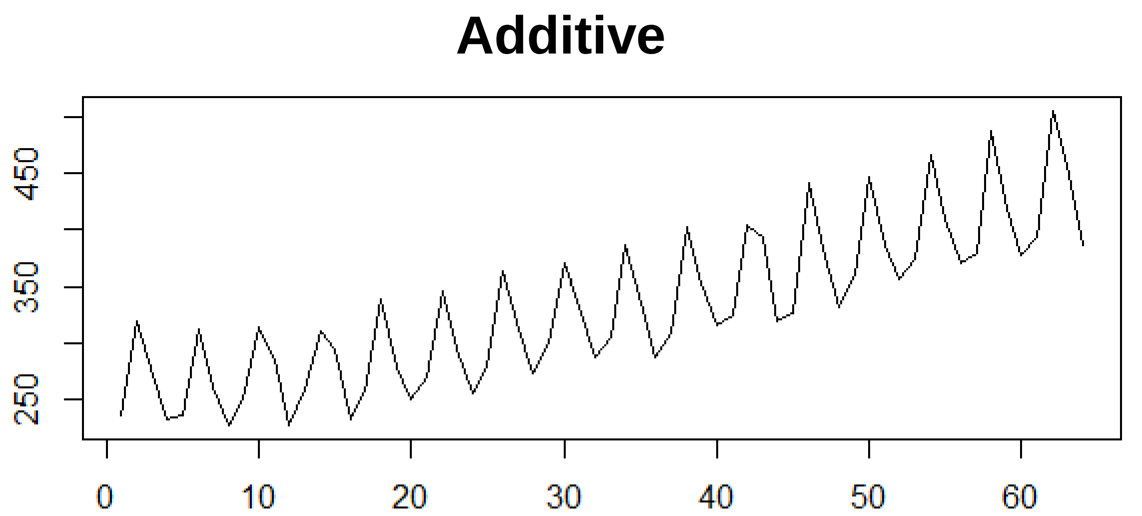
$$\boldsymbol{y_t = T_t \times S_t \times \varepsilon_t}$$

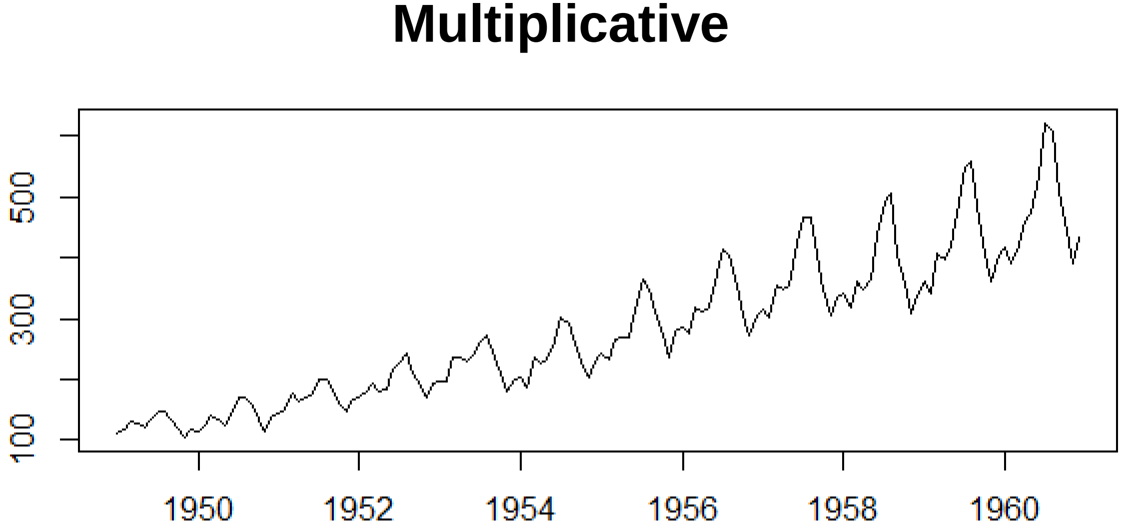
used when the variation in the seasonal pattern, or the variation around the trend-cycle, appears to be proportional to the level of the time series

(we can convert multiplicative model to additive model by taking logarithm:

$$\boldsymbol{log(y_t) = log(T_t \times S_t \times \varepsilon_t) \Rightarrow}$$
$$\boldsymbol{log(y_t) = log(T_t) + log(S_t) + log(\varepsilon_t)}$$

Additive vs. multiplicative series





How to decompose the times series ?

A simple approach to decomposition of time series would follow these steps:

1. **Estimate the trend component**

   Typically we do this by filtering time series (for instance, using moving average)

2. **Subtract (additive) or divide (multiplicative) time series by the estimated trend**

3. **Estimate seasonal component**

   Typically we calculate the means across seasons of detrended data (for instance take mean for each month across years)

4. **Calculate residuals**

   Typically we do this by subtracting (additive) or dividing (multiplicative) detrended series by seasonal component

Forecasts

- Forecast vs. predictions

- $\hat{y}_{t+k|t}$ is a forecast made at time t for a future value at time $t + k$

- Forecasts of continuous variables are typically based on <u>stochastic, linear, dynamic</u> time series models

- Primary measure of the quality of a forecast is a loss function, typically measured as the square of the forecast error

$$Loss \Rightarrow L(y_t - y_{t\,forecast}) \Rightarrow L(y_t - y_{t\,forecast})^2$$

- **Elements of a forecast:**

  Information set (data set)
  A projection date (typically the latest data for which information set is available)
  Forecast horizon (the time between projection date and predicted event)
  Model (that relates projection date to predicted event)

It can be surprisingly hard to improve forecasts of a well specified univariate time-series

Basic univariate analysis - smoothers/filters

- ***Smoothing procedure is applied with the objective of identifying an underlying signal or trend. In this setting a researcher is typically interested less in modeling than in sheer prediction.***

- *Smoothers have low costs in terms of time and effort, and computation, but they are expensive in terms of sophistication, flexibility and accuracy*

- **A lot of available options:**
  - Mean - the simplest smoother
  - Moving average
  - Hanning smoother
  - Hodrick-Prescott filter
  - Exponentially weighted moving average

- **More on moving average:**
  - Weighted
  - Loss of observations
  - Lags and leads
  - Forecast

- Typical moving average does not produce a formula which can be extrapolated for forecasting, but types of MA can be used for forecast

Exponentially weighted moving average smoother - EWMA

- Type of lagging moving average – a special case of Holt-Winters smoother
- Designed to update our estimate of a time varying mean – level
- EWMA can also account for regime change
- EWMA are backward looking so the graph is phase shifted
- EWMA infer the current mean of the series from the entire available history, but the recent observations are weighted more heavily than older observations

- Basic representation of EWMA:

$$y_t^* = \alpha y_t + (1-\alpha) y_{t-1}^*$$

- where y_t^* is the **mean** in period t, y_t is the **raw series**, α is **smoothing parameter** between 0-1. The formula is recursive, so by repeatedly substituting we get:

$$y_t^* = \alpha y_t + (1-\alpha) y_{t-1} + (1-\alpha)^2 y_{t-2} + \cdots + (1-\alpha)^i y_0^*$$

the weights are decreasing by the rate of: $\{\alpha, (1-\alpha), (1-\alpha)^2, (1-\alpha)^3, \ldots\}$

So, we end up with unknown y_0^* and α which need to be defined

Holt-Winters - generalization of EWMA

- Holt-Winters generalizes EWMA by accounting for level (mean of the series at t), slope (trend, increment between $t-1$ and t), and, optionally, seasonal component.

- There are **additive** and **multiplicative** versions of Holt-Winters. We will only focus on additive version:

$$y_t^* = a_t + b_t + s_t$$

- $a_t = \alpha(y_t - s_{t-p}) + (1-\alpha)(a_{t-1} + b_{t-1})$
- $b_t = \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1}$
- $s_t = \gamma(y_t - a_t) + (1-\gamma)s_{t-p}$

- where a_t, b_t, and s_t are the estimated **level** (mean of the series at t), **slope** (trend, increment in the series between $t-1$ and t), and **seasonal** effect at time t, and α, β and γ are the smoothing parameters. Each component of time series is modeled with separate EWMA.

Additive Holt-Winters filter

- The first EWMA ($\boldsymbol{a_t}$) takes a weighted average of our latest observation, with our existing estimate of the appropriate seasonal effect subtracted (deseasonalized).

- The second EWMA equation ($\boldsymbol{b_t}$) takes a weighted average of our previous estimate and latest estimate of the slope, which is the difference in the estimated level at time t and the estimated level at time $t - 1$.

- The third EWMA ($\boldsymbol{s_t}$), accounts for seasonal effect, by taking difference between the observation and the estimate of the level, and we take a weighted average of this and the last estimate of the seasonal effect for this season, which was made at time $t - p$ (where p is period).

Building blocks of stochastic models

- Now we are going to address some of the most important concepts of times series, in particular:
  - White noise
  - Stationarity
  - Random walk
  - Random walk with drift
  - Deterministic trend

White noise

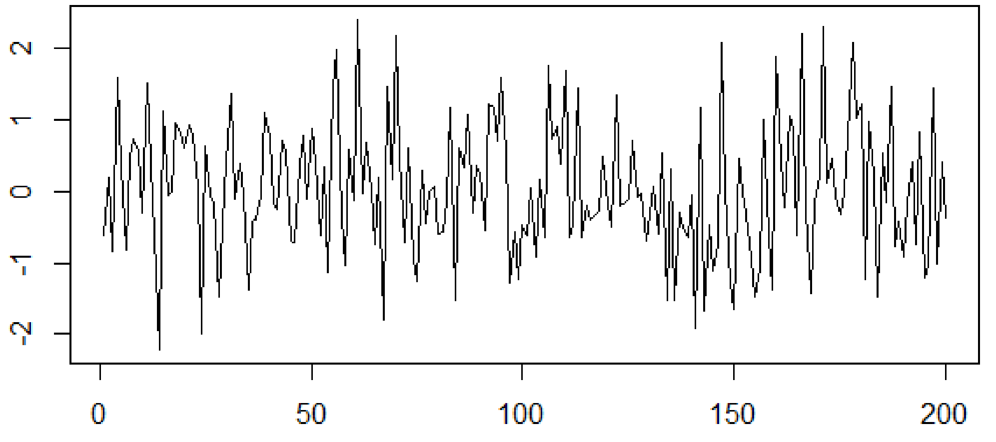
- If our model encapsulates most of the deterministic features of a time series, our residual error series should appear to be a realization of independent random variables from some probability distribution. We build models up from a model of independent random variation which is called **discrete *white noise***.

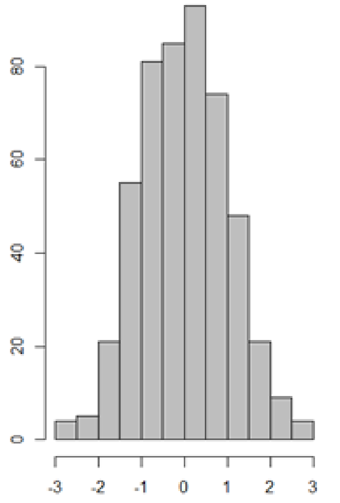
- White noise variable w_t has to satisfy:

  - $E[w_t] = 0$ mean of zero
  - $Var[w_t] = \sigma^2$ constant, finite variance
  - $Cov[w_t, w_{t+k}] = 0 \ \forall k \neq 0$ no correlation with past of future values

- Gaussian white noise has additional attribute of being normally distributed

White noise





- $\boldsymbol{y_t = \mu + w_t + \psi_1 w_{t-1} + \psi_2 w_{t-2} \ldots}$ general linear process
- General linear model converts white noise to autocorrelated series where dynamic structure of series is completely determined by $\boldsymbol{\psi}$.
- $\boldsymbol{\mu}$ is constant, but we are typically not concerned with this parameter, as it does not affects dynamics (time-dependent aspects of the model)
- White noise series is **stationary**.

Stationarity

- Stationarity means that the statistical properties of the process do not change over time.

- Two main concepts are:

➢ **Strictly stationary** – if the joint distribution of random variables of the stochastic process is independent of time. As probability distribution of time series does not depend on t, mean, variance and all higher moments are independent of t.

➢ **Weakly stationary (covariance stationary)** – if mean is constant, variance is finite and the relation between y_t and, say, y_{t-j} depends on distance between them in time, but not on their absolute location on the time line.

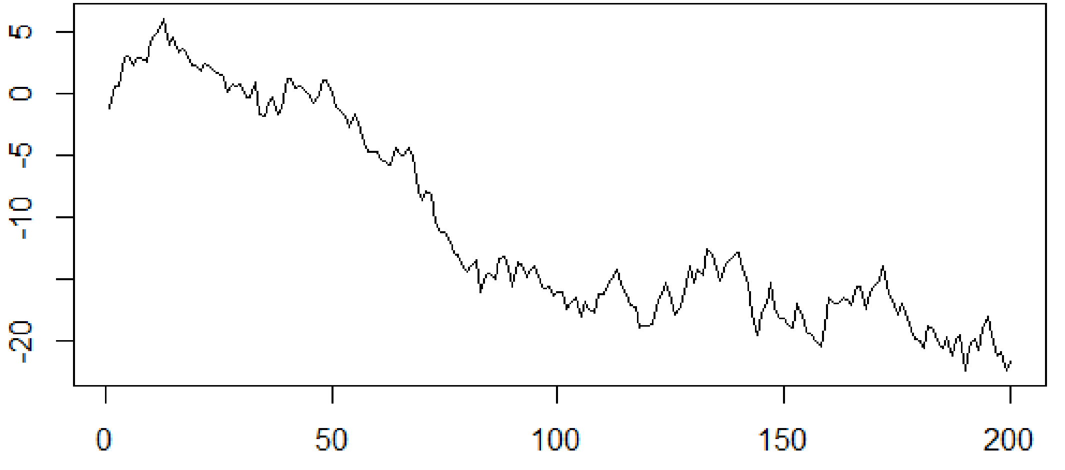
  – We typically require weakly stationary variables

Random walk (i.e. stochastic trend)

- Fundamental non-stationary model based on discrete white noise is called the **random walk**
- The series y_t is a random walk if:

$$\boldsymbol{y_t = y_{t-1} + w_t}$$

- where $\boldsymbol{w_t}$ is a white noise series. Thus, effect of $\boldsymbol{y_{t-1}}$ persists in $\boldsymbol{y_t}$



- The correlogram for a random walk is characterized by positive autocorrelations that decay very slowly down from unity – we will get back to this.

Random walk

$$y_t = y_{t-1} + w_t$$

$$y_t = y_{t-2} + w_{t-1} + w_t$$
$$y_t = y_{t-3} + w_{t-2} + w_{t-1} + w_t$$
$$y_t = y_0 + \sum_i^{t-1} w_i$$

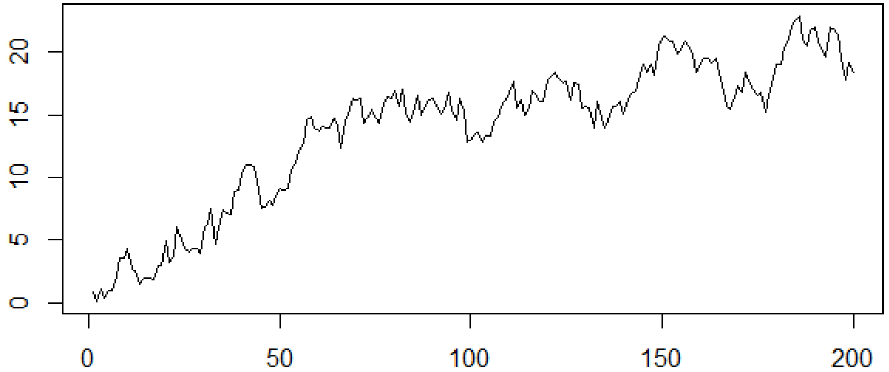
- The random walk is constant in mean. The variance increases without limit as t increases. The covariance is a function of time. So, the process is non-stationary

$$E[y_t] = 0$$
$$Cov[y_t, y_{t+k}] = \gamma_k(t)$$
$$Var[y_t] = t\sigma^2$$

Random walk with drift

- Sometimes the value of series is expected to increase despite volatility. The random walk model can be adapted to allow for this by including a drift parameter δ. Random walk with drift is defined as :

$$y_t = \delta + y_{t-1} + w_t$$
$$y_t = \delta + \delta + y_{t-2} + w_{t-1} + w_t$$
$$y_t = \delta + \delta + \delta + y_{t-3} + w_{t-2} + w_{t-1} + w_t$$
$$y_t = \delta t + y_0 + \sum_i^{t-1} w_i$$

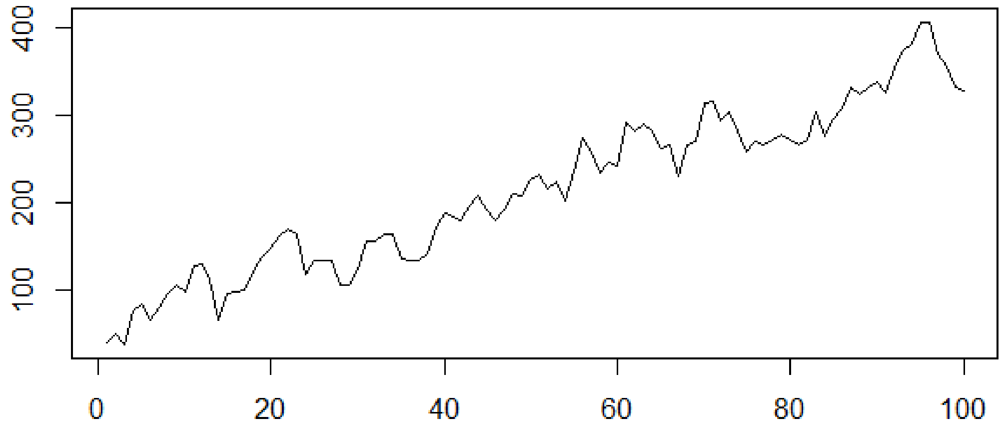


- With drift not only that the series is not stationary in the variance but it is non stationary in the mean.

$$E[y_t] = \delta t$$
$$Var[y_t] = t\sigma^2$$

Deterministic trend

$$y_t = \beta t + w_t$$



- Series is stajtionary in variance however the series is not stationary in mean. So the series is not statationary

$$E[y_t] = \beta t$$
$$Var[y_t] = \sigma^2$$

- The difference between deterministic trend and random walk with drift is that later is not stationary in variance. Even if they have same error term, deterministic trend will have less variation around the trend line in comparison to random walk with drift.

Stationary vs. non-stationary data

- Stationarity is an assumption underlying many statistical procedures used in time series analysis. However, non-stationary series can be stationary (i.e. made to be stationary) in two ways:

- **Difference stationary:** The mean trend is stochastic. Differencing adjacent terms of a series can transform a non-stationary series to a stationary series.

- **Trend stationary:** The mean trend is deterministic. Once the trend is estimated and removed from the data, the residual series is a stationary stochastic process.

(We will discuss non-stationary series in more detail later. For the rest of today's class we will mostly deal with stationary series.)

ARMA/ARIMA

(Main citation: George E.P. Box; Gwilym Jenkins (1970) Time Series Analysis: Forecasting and Control)

- A time series can be explained either by its history or by contemporaneous and past shocks. This is the goal of ARIMA - autoregressive integrated moving average model

- This method models both **stationary (AR, MA, ARMA**) and **non-stationary (ARIMA)** univariate time series

- ARIMA modeling is about modeling the correlation patterns observed in the data

- *The goal:* ***to reduce a variable to "white noise"*** *- a stochastic series. ARMA/ARIMA is designed to identify and eliminate systematic sources of error*

(later on you can focus on how white noise relates to other explanatory variables)

Notation - lag operators

- $L\varepsilon_t \equiv \varepsilon_{t-1}$
- $L(L\varepsilon_t) \equiv L(\varepsilon_{t-1}) \equiv \varepsilon_{t-2} \equiv L^2\ \varepsilon_t$

- So, we can take the general linear model:

$$y_t = \mu + \varepsilon_t + \ \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-2} \ldots$$

and represent it like:

$$y_t = \mu + \varepsilon_t + \ \psi_1 L\varepsilon_t + \psi_2 L^2\varepsilon_t$$

- By pretending that L is ordinary algebraic variable - this based on "calculus of operators" – we get:

$$y_t = \mu + (1 + \ \psi_1 L_t + \psi_2 L^2)\varepsilon_t$$

- Focusing on "polynomial part", the one in the brackets:

$$\psi(L) = 1 + \ \psi_1 L_t + \psi_2 L^2$$

$$\boldsymbol{y_t = \psi(L)\varepsilon_t}$$

ARMA – modeling stationary series

- ARMA is defined as:

$$\varphi(L)y_t = \vartheta(L)w_t$$

- where $\varphi(L)$ and $\vartheta(L)$ are finite order lag polynomials, y_t is observed time series and w_t is white noise

- The last nonzero parameter defines the order of the model:

$$(1 + \varphi_1 L + \varphi_2 L^2)y_t = (1 + \vartheta_1 L)w_t \text{ - ARMA (2,1)}$$

- To rewrite ARMA as a general linear process multiply both sides of equation $\varphi(L)y_t = \vartheta(L)w_t$ by inverse of $\varphi(L)$:

$$y_t = \varphi^{-1}(L) \times \vartheta(L)w_t$$
$$\psi(L) = \varphi^{-1}(L) \times \vartheta(L)$$
$$y_t = \psi(L)w_t$$

AR (autoregressive) process

- Linear combination of lagged <u>observed</u> variable is AR component – AR(p), ARMA(p,0) or ARIMA(p,0,0)

- A simple autoregressive model to capture a significant lag-1 autocorrelation is:

$$y_t = \mu + \varphi y_{t-1} + w_t$$

where is μ constant, w_t is withe noise error term, φ is autoregressive coefficient and $\varphi < 1$

- An AR process describes geometric (exponential) decay in the effects of shocks. As such some portion of shock lasts forever. However, practically the impact can safely be treated as zero after a few periods

- Notice: autoregressive process is a stochastic process and autocorrelation is one of the violations of the assumptions of the linear regression model (although often authors use term autocorrelation more freely).

AR process

- The model is a regression of y_t on past terms from the same series; hence the use of the term ‘autoregressive’.

- The series is an autoregressive process of order p, abbreviated to AR(p), if

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + w_t$$

or in terms of back shift operator

$$\left(1 - \varphi_1 L - \varphi_2 L^2 - \cdots - \varphi_p L^p\right) y_t = w_t$$

- The random walk is the special case of AR(1) with $\varphi = 1$:

$$y_t = y_{t-1} + w_t$$

- The exponential smoothing model is the special case:

$$\varphi_i = \varphi(1 - \varphi)^i$$

Estimation of AR process

- AR model can be fit by OLS, maximum likelihood, or using Yule–Walker equations.

- Yule Walker equations relate the AR model parameters to the autocovarinace (γ) of the random process

$$\begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \vdots \\ \gamma_p \end{bmatrix} = \begin{bmatrix} \gamma_0 & \gamma_{-1} & \gamma_{-2} & \cdots \\ \gamma_1 & \gamma_0 & \gamma_{-1} & \cdots \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \\ \gamma_{p-1} & \gamma_{p-2} & \gamma_{p-3} & \cdots \end{bmatrix} \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \varphi_3 \\ \vdots \\ \varphi_p \end{bmatrix}$$

- However, Yule-Walker equations and other estimation methods do not address the problem of autoregressive order (how much history a model should have, ***p*** order of AR(p)).

MA (moving average) process

- Linear combination of unobservable white noise disturbances is MA process - MA(q), ARMA(0,q) or ARIMA(0,0,q)

- A simple moving average model MA(1) is defined as:

$$y_t = \mu + w_t + \vartheta w_{t-1}$$

where is μ constant, w_t is withe noise error term,ϑ is moving average coefficient

- Moving average of order q is defined as :

$$y_t = w_t + \ \vartheta_1 w_{t-1} + \vartheta_2 w_{t-2} + \cdots + \vartheta_q w_{t-q}$$
$$y_t = \left(1 + \vartheta_1 L + \vartheta_2 L^2 + \cdots + \vartheta_q L^q\right) w_t$$

- Unlike AR process where autocorrelation function has a long tail, MA process rapidly decays once you pass the history.

Estimation of MA process

- The parameters of MA model cannot be estimated by OLS because:

    - random shock are not given
    - solving for error terms is dependent on estimated parameters

- A numerical fitting method tries all coefficient sets on a finite grid of parameter values and selects the one which minimizes sum of the random shocks.

Infinite AR and MA representations

AR(1) to MA(∞)

- $y_t = \varphi y_{t-1} + w_t$
- $y_t - \varphi y_{t-1} = w_t$
- $(1 - \varphi L) y_t = w_t$
- $y_t = \frac{w_t}{(1-\varphi L)}$

Assuming that modulus of |φ|<1, it can be proved that some infinite sum:

$\alpha + \alpha\varphi^2 + \alpha\varphi^3 + \alpha\varphi^4 + \ldots = \frac{\alpha}{1-\varphi}$

- $y_t = w_t + \varphi L w_t + \varphi L^2 w_t + \cdots$
- $y_t = w_t + \varphi w_{t-1} + \varphi^2 w_{t-2} + \cdots$

MA(1) to AR(∞)

- $y_t = w_t + \vartheta w_{t-1}$
- $y_t = (1 - \vartheta L) w_t$
- $\frac{y_t}{(1-\vartheta L)} = w_t$

If modulus |θ|<1:

- $y_t + \vartheta L y_t + \vartheta^2 L^2 y_t + \cdots = w_t$
- $y_t + \vartheta y_{t-1} + \vartheta^2 y_{t-2} + \cdots = w_t$
- $y_t = w_t - \vartheta y_{t-1} - \vartheta^2 y_{t-2} - \cdots$

- **If AR(p) process can be represented as an infinite MA process the series satisfies <u>stability condition</u>**
- **If MA(p) process can be represented as an infinite AR process the series satisfies <u>invertibility condition</u>**

Dynamic characteristics of AR and MA

- We typically address dynamics of the system by analyzing impact of single non-zero w_t on y_t. In other words, we set $w_0 = 1$ and other $w_t = 0$, and the coefficients trace the impact of this single random impulse – this is called **impulse response function**

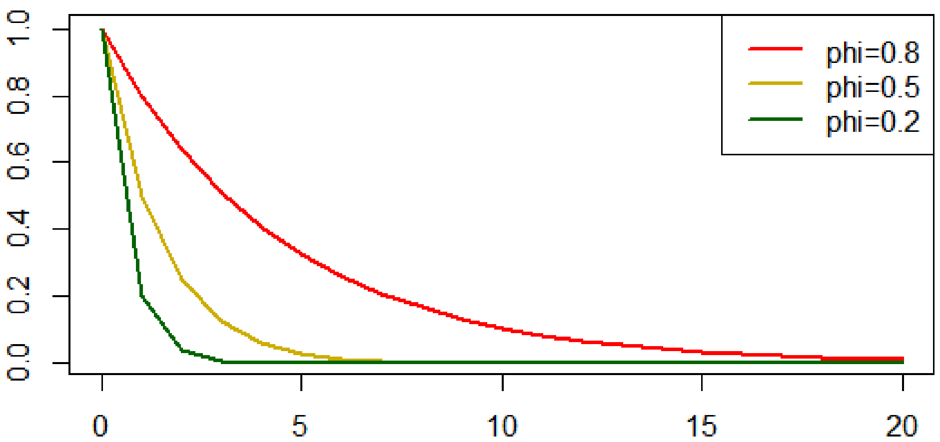
- In simplest case of the AR(1) process, the future values of y are given by:

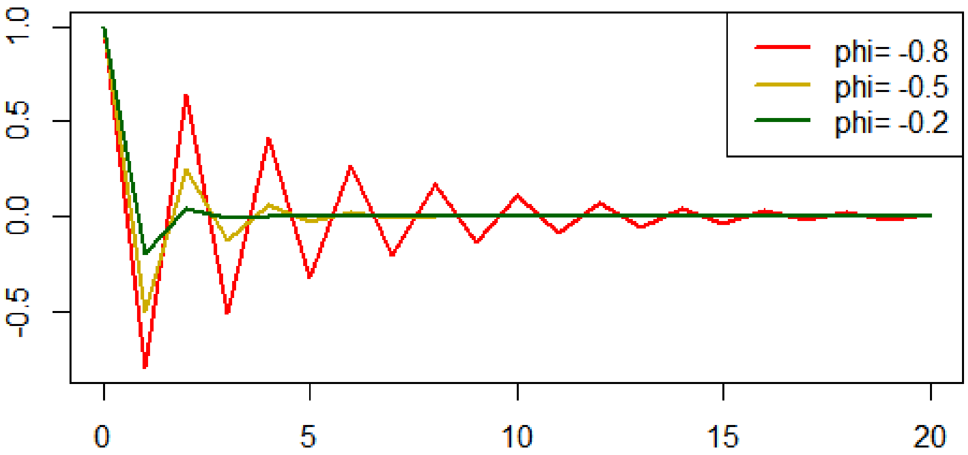
$$y_{t+j} = \varphi_1^j y_t$$

where values exponentially decay towards 0

Dynamic characteristics of AR(1)

- Dynamic response to a unit disturbance in AR(1) with positive and negative values of φ





Dynamic characteristics of MA(q)

- In a pure MA model, if the final coefficient is ϑ_{t-k} than a single shock has no impact after k periods. So, in comparison to AR, MA component is less persistent.

- However, when the model contains an AR component, the impact of single shock persists forever although magnitude of the impact decays over time

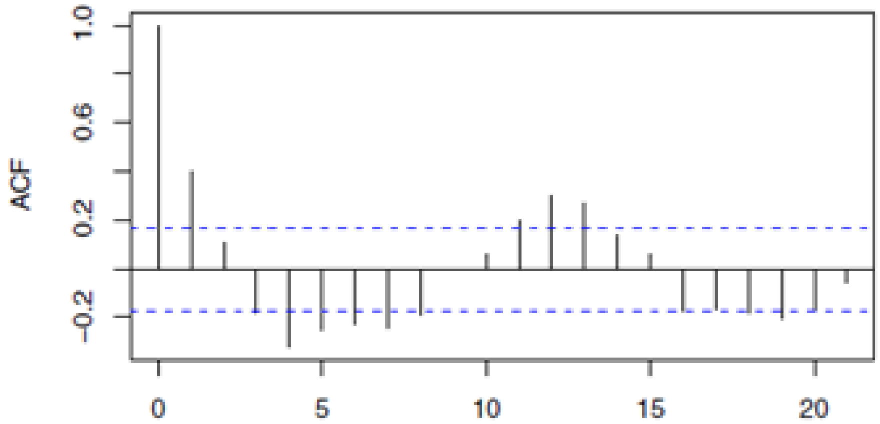
Correlograms

- **Discovering structure of correlation of a variable with itself at different times is crucial in understanding order of ARMA(p,q).**

- **The autocorrelation function** (acf) accounts for the correlation of series with itself at lag k, where from the definition ρ_0 is 1. A plot of the estimated autocorrelations can be useful when determining the order of a suitable MA process.

- **The partial autocorrelation** (pacf) at lag k is the correlation that results after removing the effect of any correlations due to the terms at shorter lags. A plot of the estimated partial autocorrelations can be useful when determining the order of a suitable AR process.

Correlograms



- If a spike falls outside dotted lines, we have evidence against the null hypothesis that $\rho_k = 0$ at the 5% level. However, these estimates are correlated, so if one falls outside the lines, the neighboring ones are more likely to be statistically significant. Also sampling variation may influence significance.

- Usually a trend in the data will show in the correlogram as a slow decay in the autocorrelations, which are large and positive.

- If there is seasonal variation, seasonal spikes will be superimposed on the pattern in correlogram.