Time Series Analysis & Forecasting

Class 1

Arnab Bose, Ph.D.

MSc Analytics

University of Chicago

Time Series (TS)

uniform: at reg time intervals

- Data is measured at different time points, uniform and non-uniform (uneven spacing)
- Primary objective of TS analysis is to develop mathematical models that provide plausible description of the data
- Data collection -> separate field of study that outlines the challenges in data collection

when it is uniform, the math is much more simple, so this is what we are going to focus on in class unless otherwise specified

we want to create mathematical models that describe data generation process

Random Variable (RV)

- A variable whose values are defined by a probability distribution. The distribution specifies the probability that the value of the variable is within any given interval.
- RV can be

  discrete – taking finite or countable list of values

  continuous – taking any numerical value

White Noise

- Collection of uncorrelated random variables w_t that are uncorrelated with mean 0 and variance σ_w^2
- White – analogy with white light where all possible frequencies are present with equal strength
- Gaussian white noise is independent and identically distributed (iid)

Random Walk

- Mathematically represented as $Y_t = Y_{t-1} + \, w_t$ TS {Yt} = Yt-2, Yt-1, Yt, Yt+1, ….
- Current value Y_t is previous value Y_{t-1} plus completely random white noise w_t.

- Autocovariance – measures the linear dependency between two points in the same TS observed at different times. the values in a time series depend on each other, so high value, smoother line

Brownian Motion: molecule follows this
Stock prices are supposed to have a random walk
Random walk building upon white noise

auto = self
co = together
variance = deviance from mean
lower case gamma, kinda looks like a V
Vs,t = E {(Xs - ux)(Xt-ux)]
Vk = ‘ ‘

autocovar = cov at 2 diff lags / sqrt(cov of x at first lag times cov x at second lag)
so bottom is var (x lag 1) * var (x lag 2)
the bottom is the covariance of the of itself with no lag which is var

correlation is the normalized version of covariance
Corr(xt,yx) = Cov (x,y) / Cov(x)*Cov(y)

autocorrelation at lag 0 equal 1 since you are completely correlated with yourself

Stationarity

hard to detect from a graph

the data generation process is in statistical equilibrium

- TS is said to be stationary – the process is in statistical equilibrium.
- Basic idea is that the laws of probability that govern the process behavior do not change over time.

probability laws that govern data generation processes do not change with time

two ways to define stationary

- STRICTLY stationary – when the probability behaviors of every collection of values of a TS $\{x_t\}$ is identical to that of the time shifted TS $\{x_{t+h}\}$ for all h

the prob laws remain unchanged throughout time
the joint probability distribution of the TS is constant over time
each random variable (at each t) — look at green graph — is just a number from prob dist

- WEAKLY stationary - $\{x_t\}$ is a finite variance process such that

= wide sense stationary

  - Mean is constant and does not depend on t
  - Autocovariance function $\gamma(t, s)$ depends on t, s through their difference $|t - s|$

autocorvariance (lowercase gamma) only dependent on lag
autocor is constant for a lag, so between mon and wed is same as fri and sun
if we have a joint prob dist (which is product of each marginal distribution), and does not depend on time => STRICT
i.e. marginal distributions are iid then STRICT
cant say the other way since we can't say its independent (essentially the RV all have the same marg dist and joint dist)

way to test independence, if you change a value at t1, see if following values change

- If TS is Gaussian then weak stationarity === strict stationarity

if we have proven its weak, then we know it is strict

iid = identical, and independently distributed

if we have 2 RV that are iid, we know they are uncorrelated

cauchy variance has undefined var, so it can not be weakly stationary

order of moments:
1. mean
2. variance
3. skew
4. kertosis

white noise is weak stationary
random walk is not stationary

sin(t) is deterministic, we know what it will be in the future
we are interested in stochastic processes

Random Walk

not stationary because autocov = tσ2w
autocov changes with time

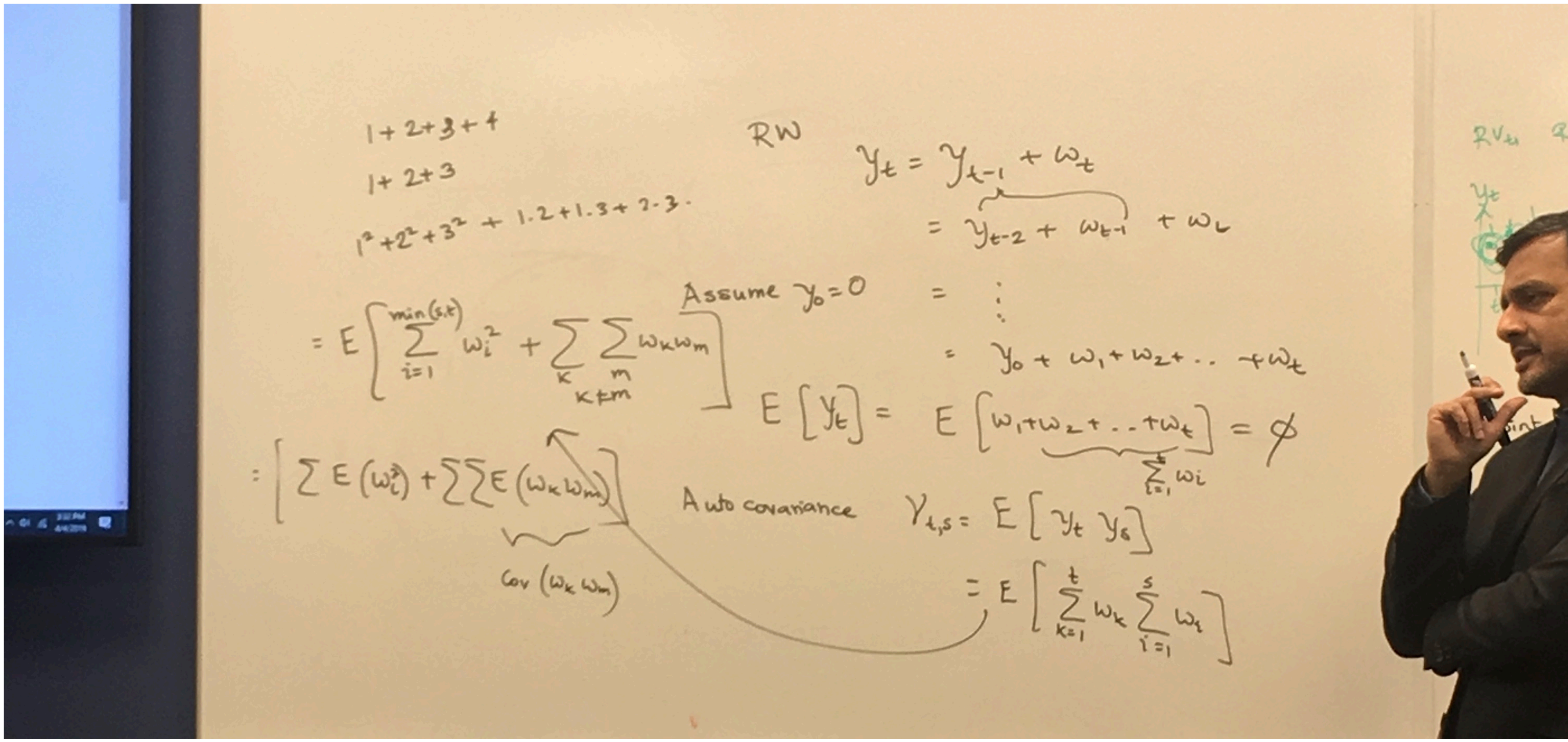
RW: Yt = Yt-1 + wt
can expand
Yt-1 = Yt-2 + wt-1

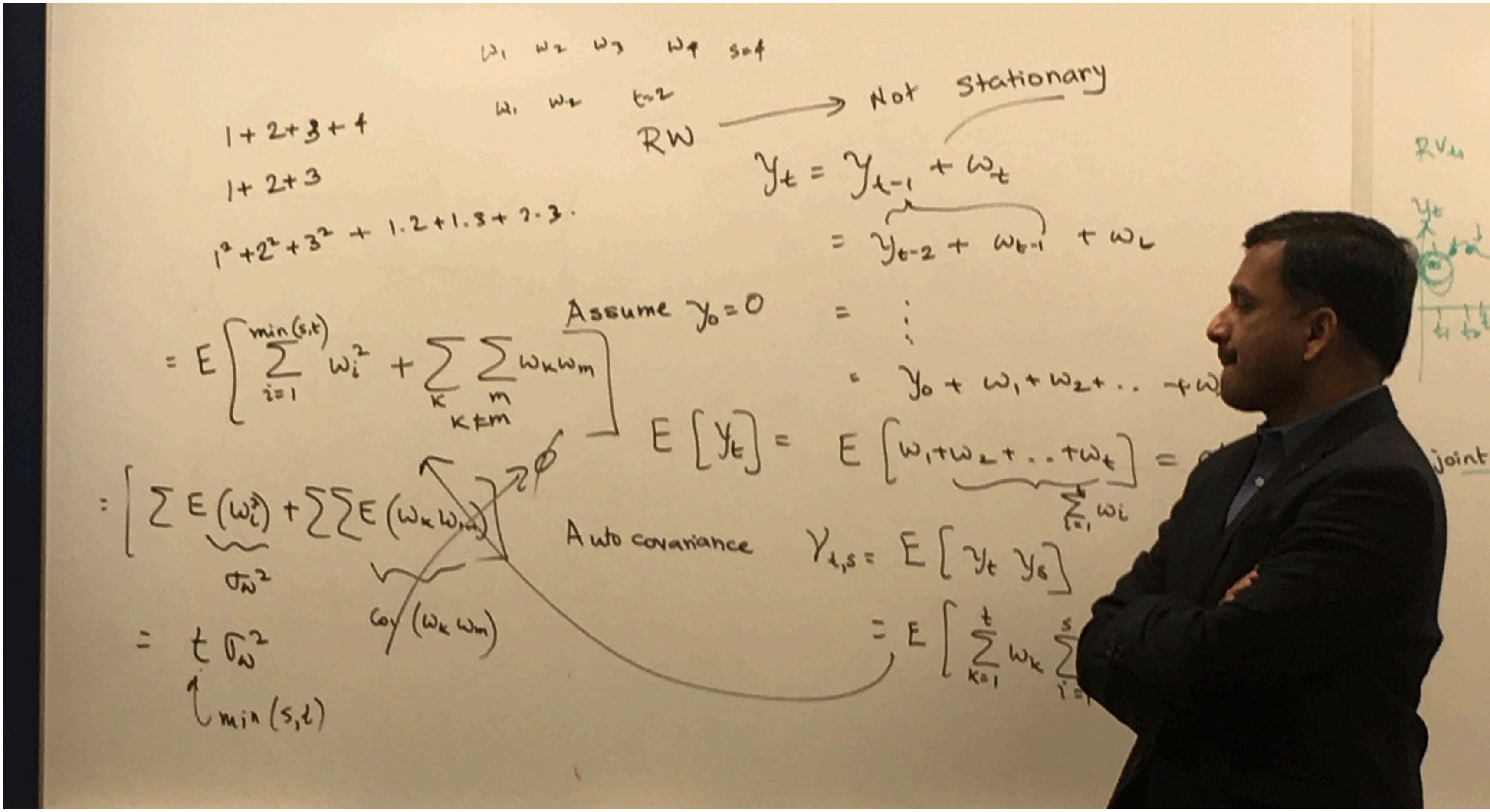
assume Y0 = 0

E[Yt] = E[w1 +w2 + .. + wk] = 0

- Mean $\mu = 0$
- Variance $\gamma(0) = t\sigma_w^2$
- Autocovariance $\gamma_{t,s} = \min\{t, s\}\ \sigma_w^2$
- Autocorrelation $\rho_{t,s} = \sqrt{\frac{t}{s}}, \forall\ t \leq s$

uncorrelated RV, so autocovariance is 0


$1+2+3+4$
$1+2+3$
$1^2+2^2+3^2 + 1.2+1.3+2.3.$
RW
$y_t = y_{t-1} + w_t$
$= y_{t-2} + w_{t-1} + w_t$
Assume $y_0 = 0$
$= \vdots$
$= y_0 + w_1 + w_2 + \dots + w_t$
$= E\left[\sum_{i=1}^{\min(s,t)} w_i^2 + \sum_{k}\sum_{m} w_k w_m\right]$ $k \neq m$
$E[y_t] = E[\underbrace{w_1 + w_2 + \dots + w_t}_{\sum_{i=1}^{t} w_i}] = \phi$
$= \left[\sum E(w_i^2) + \sum\sum E(w_k w_m)\right]$
Auto covariance $\gamma_{t,s} = E[y_t\, y_s]$
$= E\left[\sum_{k=1}^{t} w_k \sum_{i=1}^{s} w_i\right]$
$Cov(w_k w_m)$

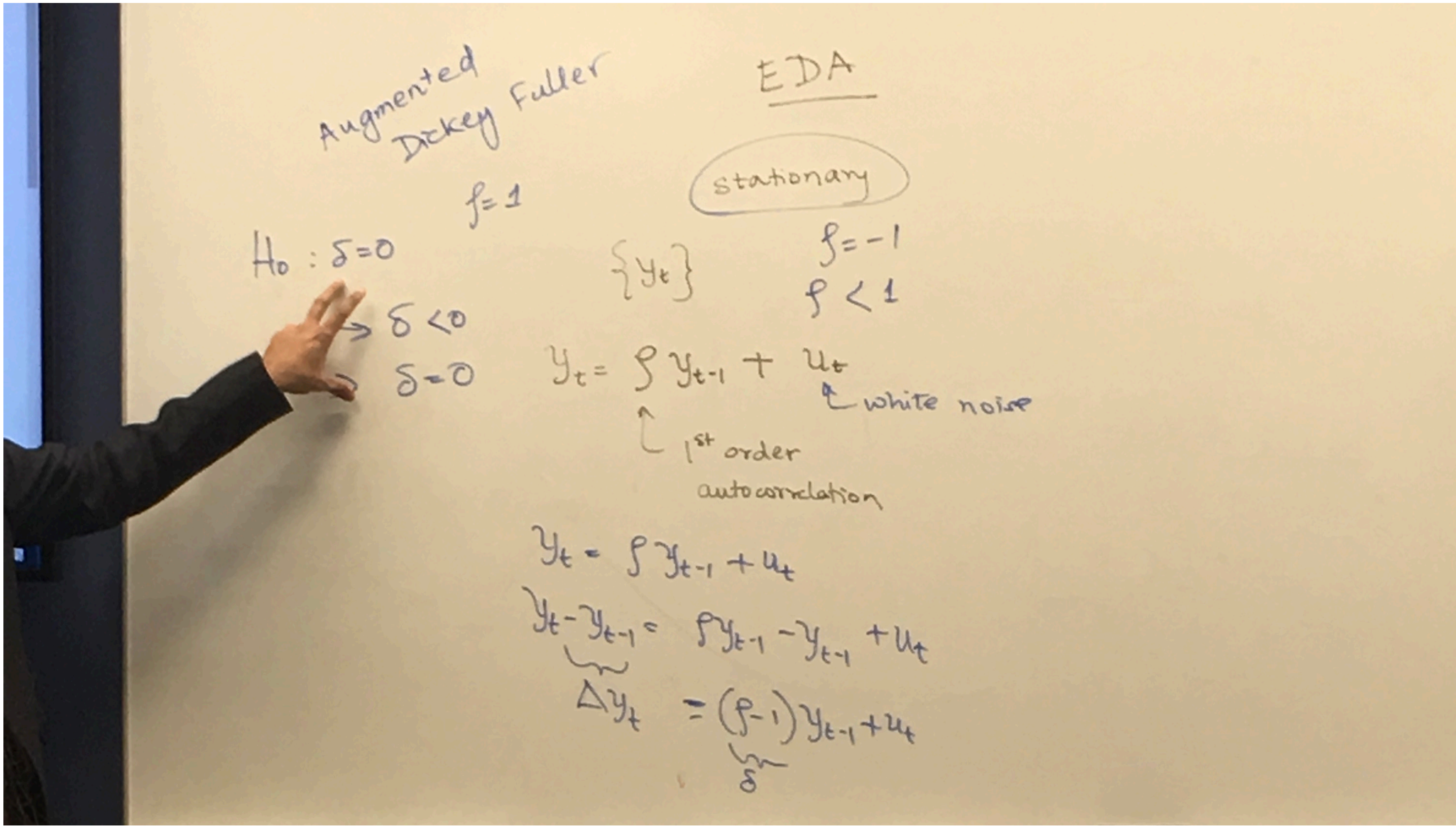

w1 w2 w3 w4 s=4
w1 w2 t=2
1+2+3+4
1+2+3
1^2+2^2+3^2 + 1.2+1.3+2.3.
RW → Not stationary
yt = yt-1 + wt
= yt-2 + wt-1 + wt
Assume y0=0
= y0 + w1 + w2 + .. + w
= E [Σ_{i=1}^{min(s,t)} wi^2 + Σ_k Σ_m wk wm] k≠m
E[yt] = E[w1+w2+..+wt] = Σ_{i=1} wi
= [Σ E(wi^2) + ΣΣ E(wk wm)]
σw^2
cov(wk wm)
Auto covariance γt,s = E[yt ys]
= E[Σ_{k=1}^{t} wk Σ_{i=1}^{s}
= t σw^2
min(s,t)
joint

Exploratory Data Analysis

1. Idea is to understand datasets, both univariate and multivariate
2. Univariate – test for stationarity
3. Consider a stochastic TS given by $Y_t = \rho Y_{t-1} + u_t$
4. To determine stationarity, use a unit root test such as Augmented Dickey Fuller (ADF)

   augmented: removes the assumption of error term being uncorrelated
5. Augmented – no assumption about the error term u_t being uncorrelated to itself
6. To transform a non-stationary TS to stationary TS, iteratively take the difference of the non-stationary TS and verify if the differenced TS is stationary
7. If TS has a unit root, then the first difference is stationary

KPSS is another test for stationarity
Ho: stationary
Ha: non-stationary
if KPSS and ADF disagree, go with KPSS because it is more stable


Augmented Dickey Fuller
EDA
stationary
$\rho = 1$
$H_0 : \delta = 0$
$\{y_t\}$
$\rho = -1$
$\rho < 1$
$\rightarrow \delta < 0$
$\rightarrow \delta = 0$
$y_t = \rho y_{t-1} + u_t$
white noise
1st order autocorrelation
$y_t = \rho y_{t-1} + u_t$
$y_t - y_{t-1} = \rho y_{t-1} - y_{t-1} + u_t$
$\Delta y_t = \underbrace{(\rho - 1)}_{\delta} y_{t-1} + u_t$

R code – white noise + random walk

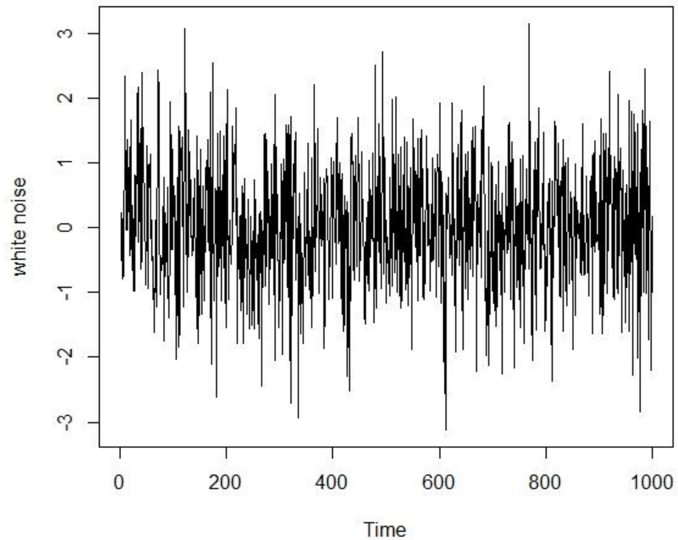
- wn <- rnorm(500, 0, 1) # White Noise
- rw <- cumsum(wn) # Random Walk

- plot.ts(wn, main="white noise")
- plot.ts(rw, main="random walk")

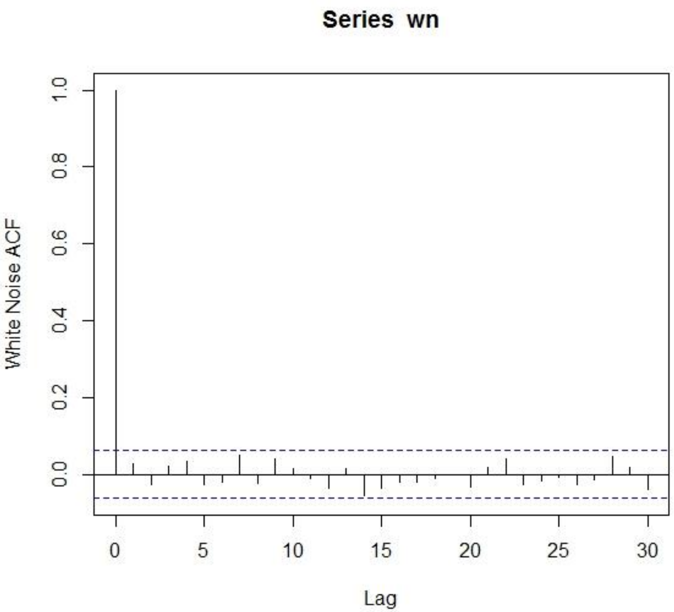
- acf(wn) **#0 lag of will always be 1, the blue lines are 95% confidence interval, rest is not correlated**
- acf(rw)

- library("tseries")
- adf.test(wn) **expect low p since white noise is stationary, and Ho: non-stationary** **kpss.test(wn) - expect high value**
- adf.test(rw) **expect large value for p, to show that it is nonstationary** **kpss.test(rw) - expect a low value**

R plot – white noise

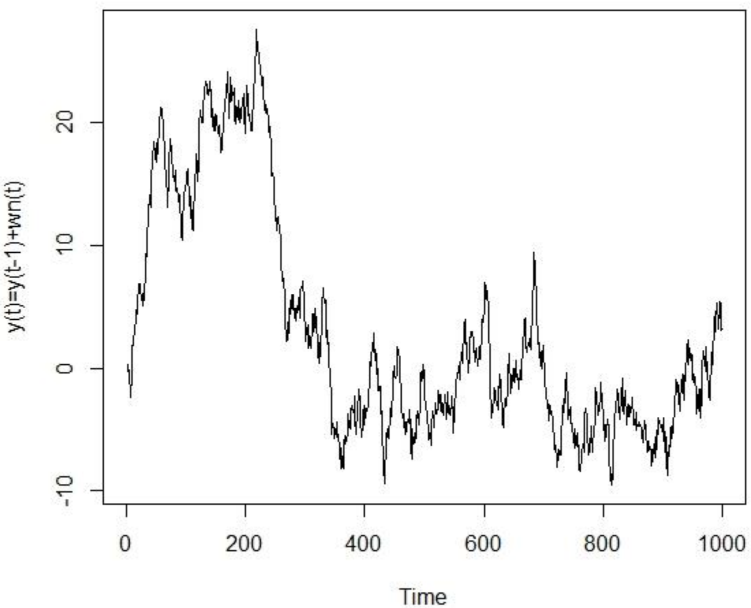


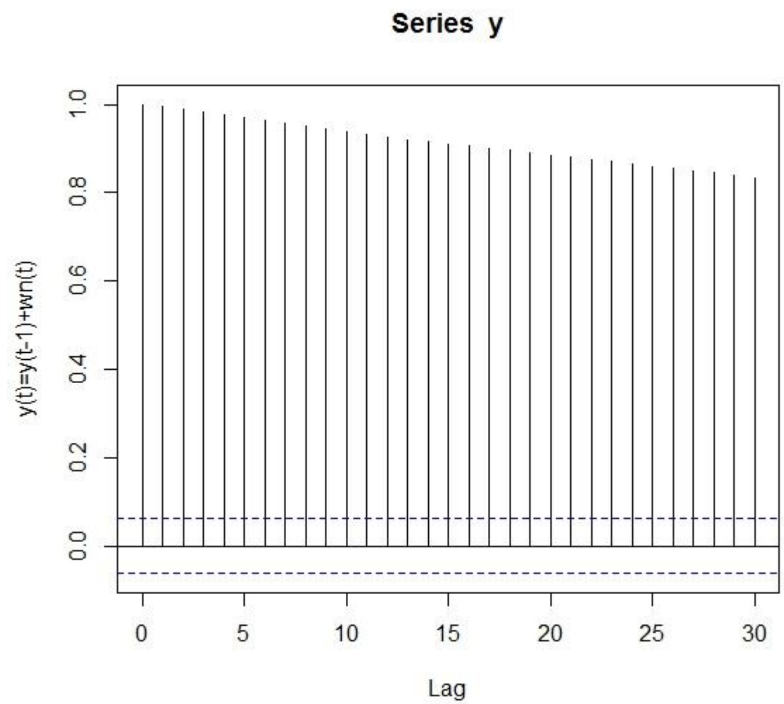
not guaranteed, but a good way to see if its stationary



plot is qualitative, doing KPPS and ADF is the quantitative way to detect stationarity

R plot – random walk



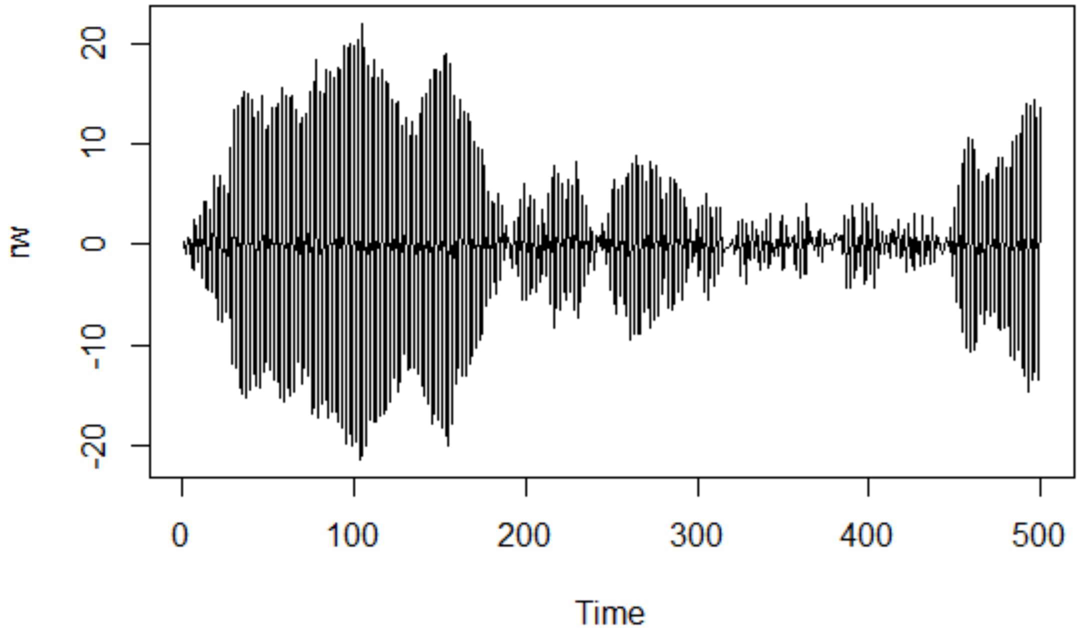


R code – random walk with $\boldsymbol{\rho = -1}$

- rw <- c(0)
- wn <- rnorm(500, 0, 1)
- rw[1] <- wn[1]
- for (i in 2:length(wn))
- rw[i] <- -1 * rw[i-1] + wn[i]
- ts.plot(rw)
- adf.test(rw)
- kpss.test(rw)

Random walk with $\rho = -1$

change to- .3, looks less symetrical
change to -.95 it is closer to symetry



Textbook Chapters and Assignment

- Materials covered available in book:
  - FPP: Chapters 1 – 3, MJK: Chapters 1 – 2, TSA: Chapters 1 – 2