"Success is not final; failure is not fatal: It is the courage to continue that counts." — Winston S. Churchill

ARMA

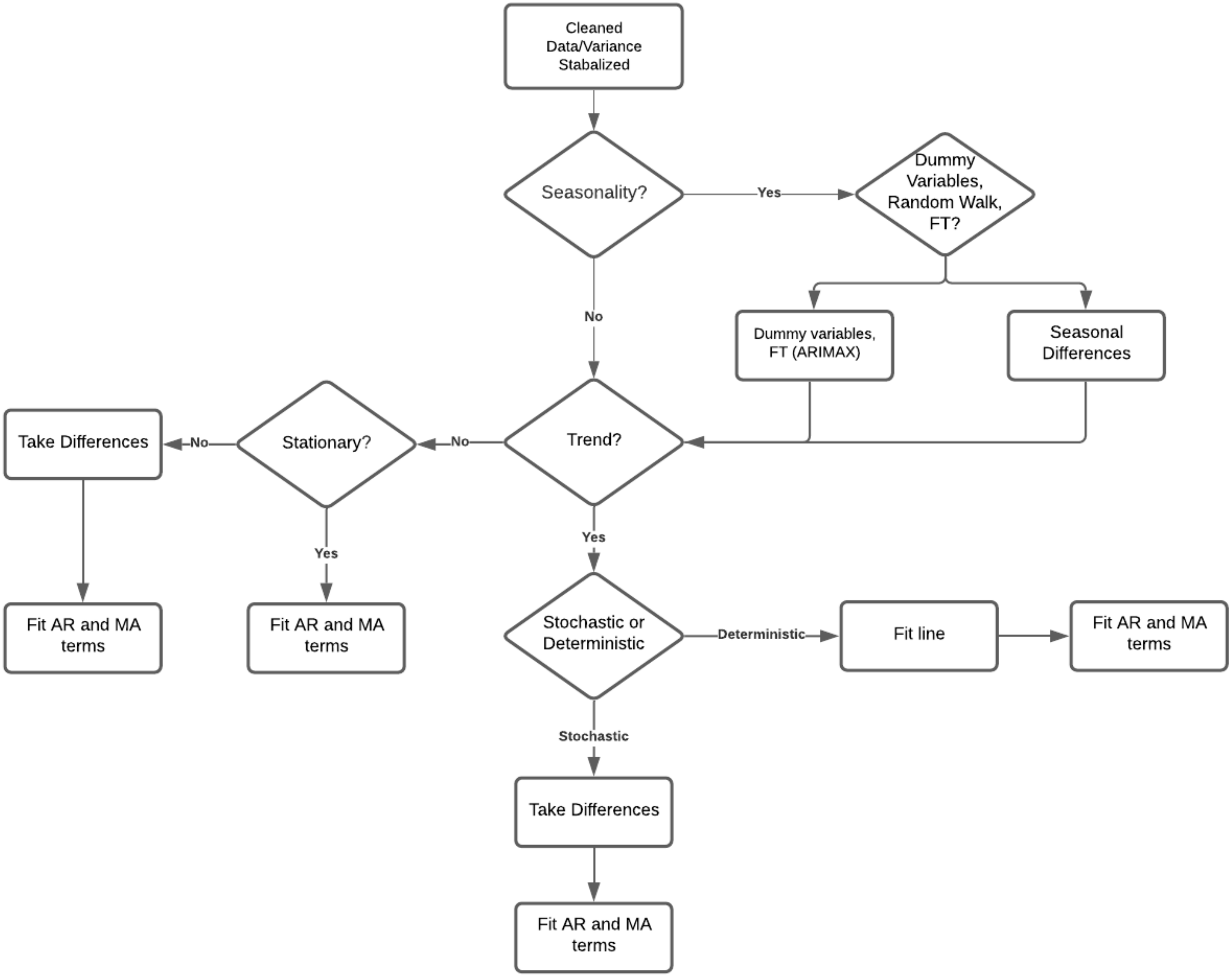
ARMA

- ARMA stands for AutoRegressive Moving Averages (AR and MA terms are used to model the dependency structure in the data!)
- ARMA models are based upon statistical methods (will assume a distribution!!)
- When creating ARMA models, it can be a circular process (when changing something later in a model might make you reevaluate what you did earlier)
- Best model will be found by an iterative process!!

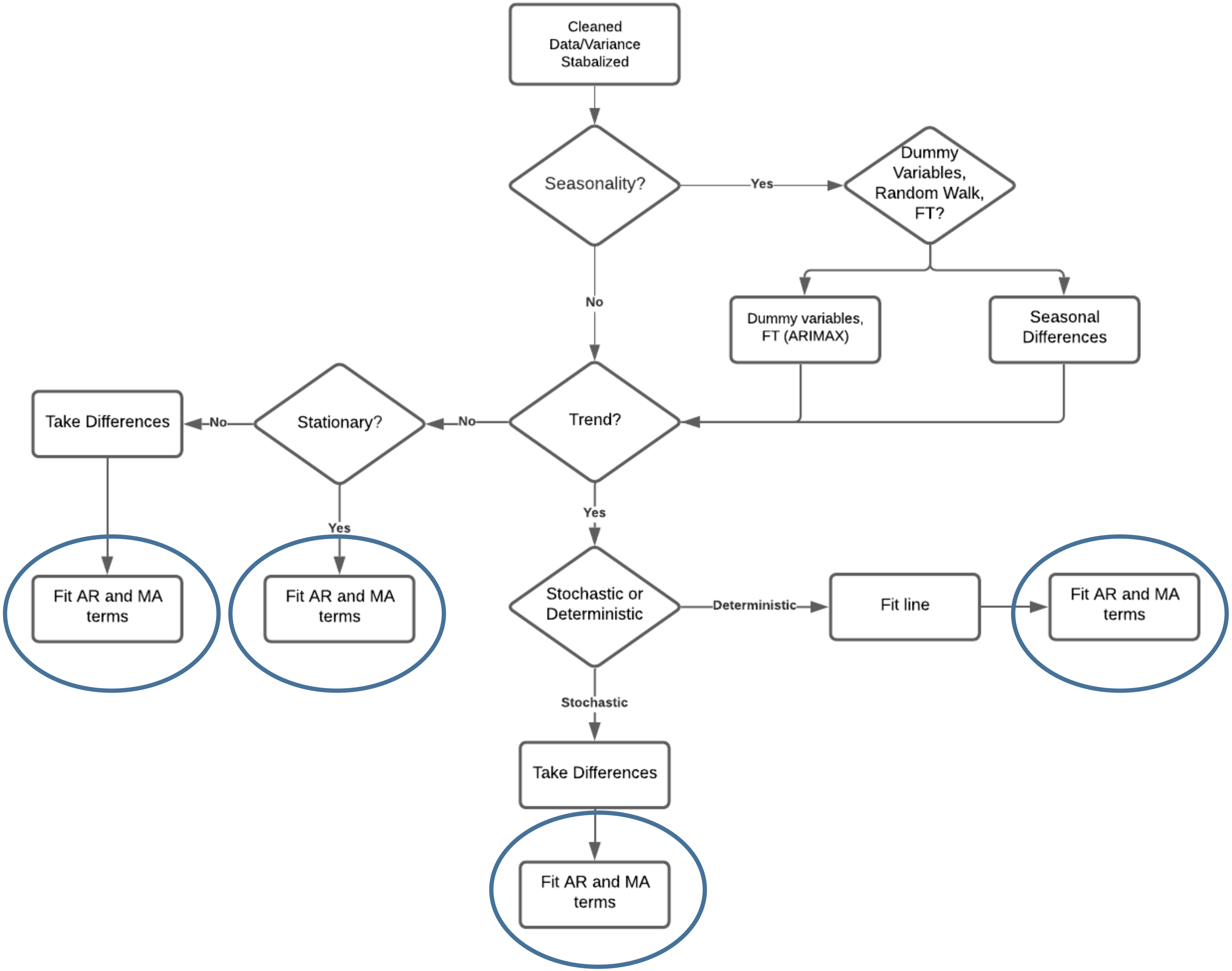
ARMA

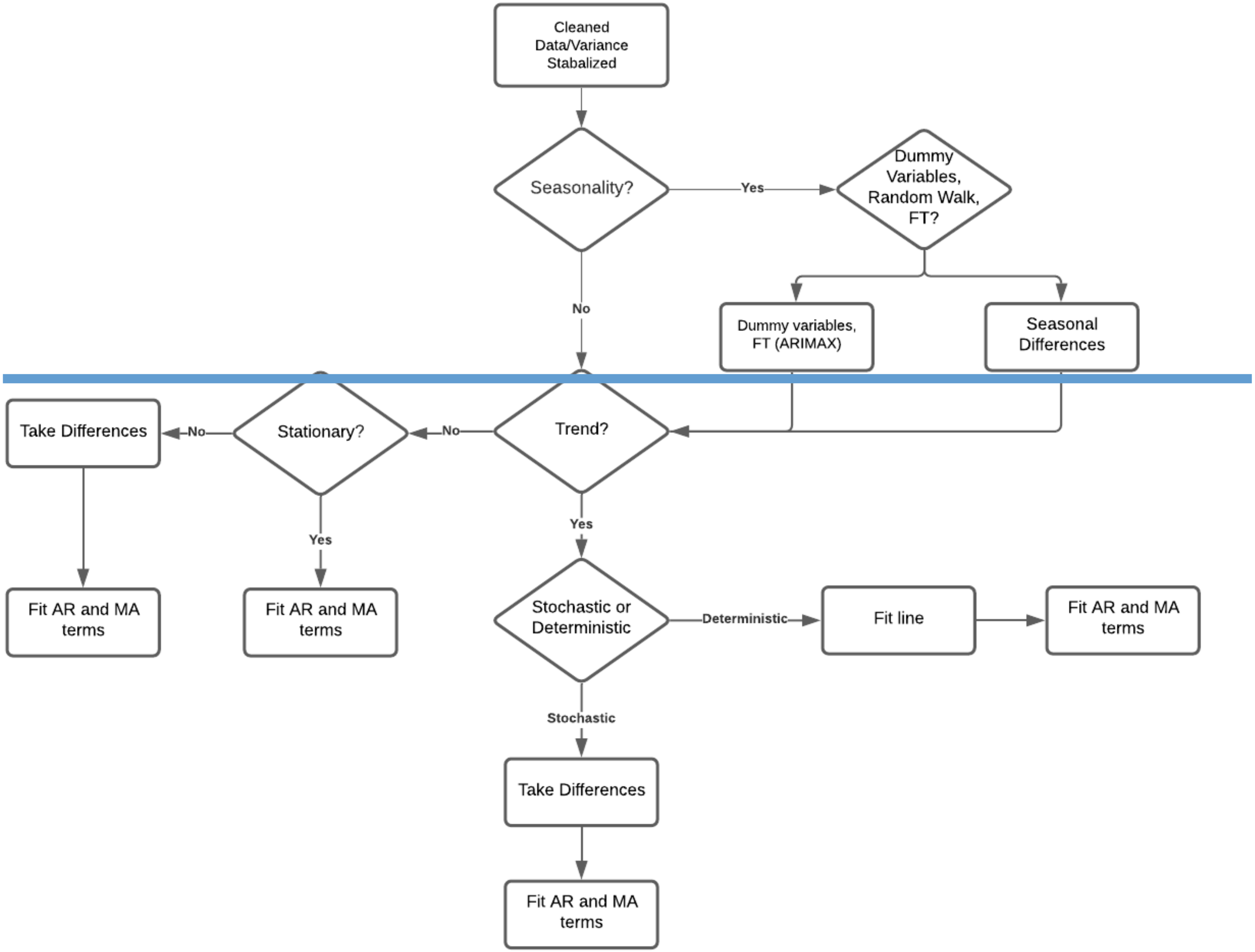
SIGNAL:

- Can have signal due to a seasonal pattern (will be discussed in Fall 2)
- Can have signal due to trend (we will discuss later in this class)
- Can have signal due to “correlation structure” which can be in the form of Autoregressive (AR) and moving averages (MA)
  - However, in order to model the dependency in the data appropriately, we will need to take care of the ***functional form*** (for example trend and/or seasonality) and any random walks first
  - After accounting for the above information, we can then begin modeling the dependency structure in the data



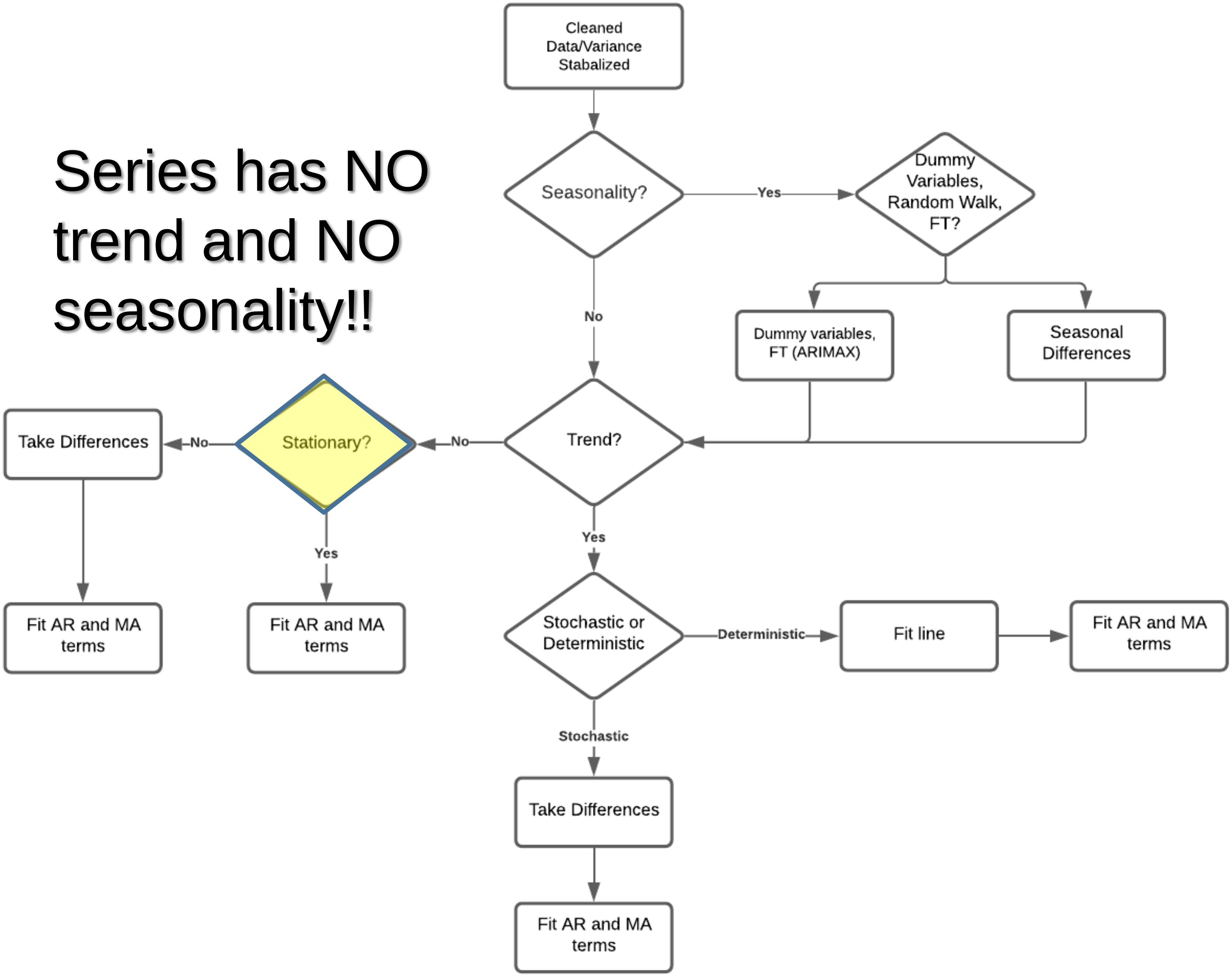
Cleaned Data/Variance Stabalized
Seasonality?
Yes
Dummy Variables, Random Walk, FT?
No
Dummy variables, FT (ARIMAX)
Seasonal Differences
Take Differences
No
Stationary?
No
Trend?
Yes
Yes
Fit AR and MA terms
Fit AR and MA terms
Stochastic or Deterministic
Deterministic
Fit line
Fit AR and MA terms
Stochastic
Take Differences
Fit AR and MA terms


Cleaned Data/Variance Stabalized
Seasonality?
Yes
Dummy Variables, Random Walk, FT?
Dummy variables, FT (ARIMAX)
Seasonal Differences
No
Trend?
No
Stationary?
No
Take Differences
Yes
Fit AR and MA terms
Fit AR and MA terms
Yes
Stochastic or Deterministic
Deterministic
Fit line
Fit AR and MA terms
Stochastic
Take Differences
Fit AR and MA terms


Cleaned Data/Variance Stabalized
Seasonality?
Yes
Dummy Variables, Random Walk, FT?
No
Dummy variables, FT (ARIMAX)
Seasonal Differences
Take Differences
No
Stationary?
No
Trend?
Yes
Yes
Fit AR and MA terms
Fit AR and MA terms
Stochastic or Deterministic
Deterministic
Fit line
Fit AR and MA terms
Stochastic
Take Differences
Fit AR and MA terms

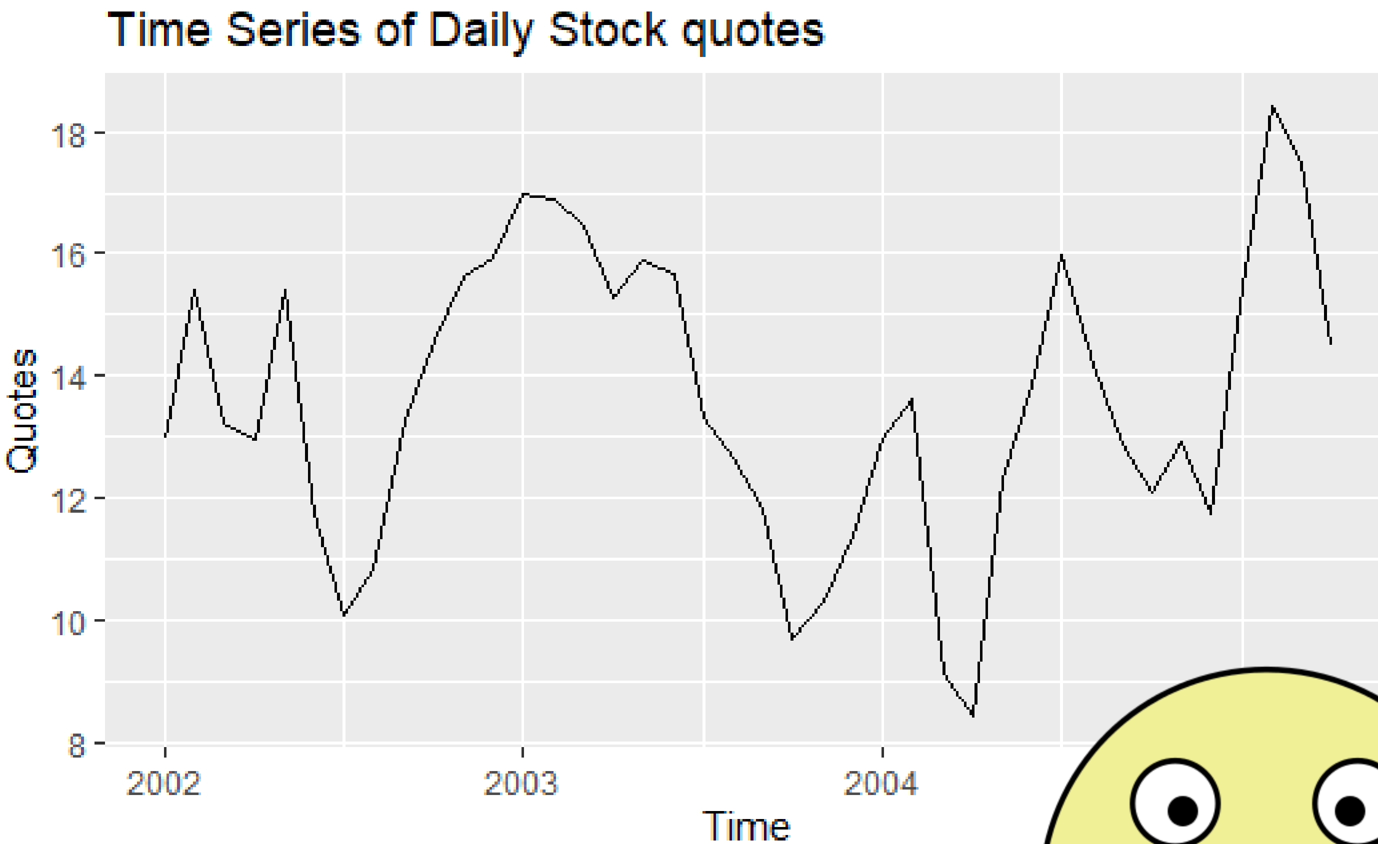
NO SEASON AND NO TREND (START SIMPLE….)

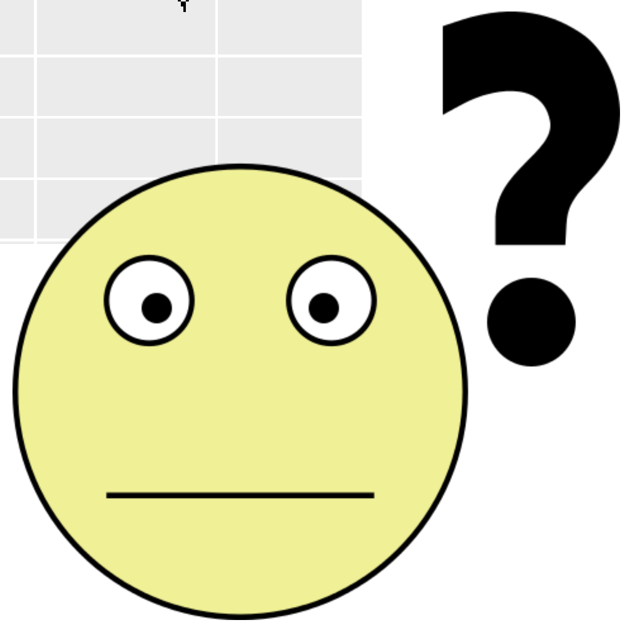
Series has NO trend and NO seasonality!!



Stationarity

- To model the AR and MA terms, we ***must*** have stationarity first
- We will be using the idea of “weak stationarity” for modeling
  - No predictable pattern, constant variance and converges to a constant mean in the “long run”


Time Series of Daily Stock quotes
Quotes
18
16
14
12
10
8
2002
2003
2004
Time



What is a ‘Random Walk’?

Random Walk Model

- Random Walk Model:

$$Y_t = Y_{t-1} + e_t$$

Random Walk Model

- There are two types of random walk models (random walk with drift and random walk without drift):

$$Y_t = Y_{t-1} + e_t$$

Best guess for Y_t is Y_{t-1}.
Best guess for Y_{t-1} is Y_{t-2}…etc

Stochastic Trend: Differencing

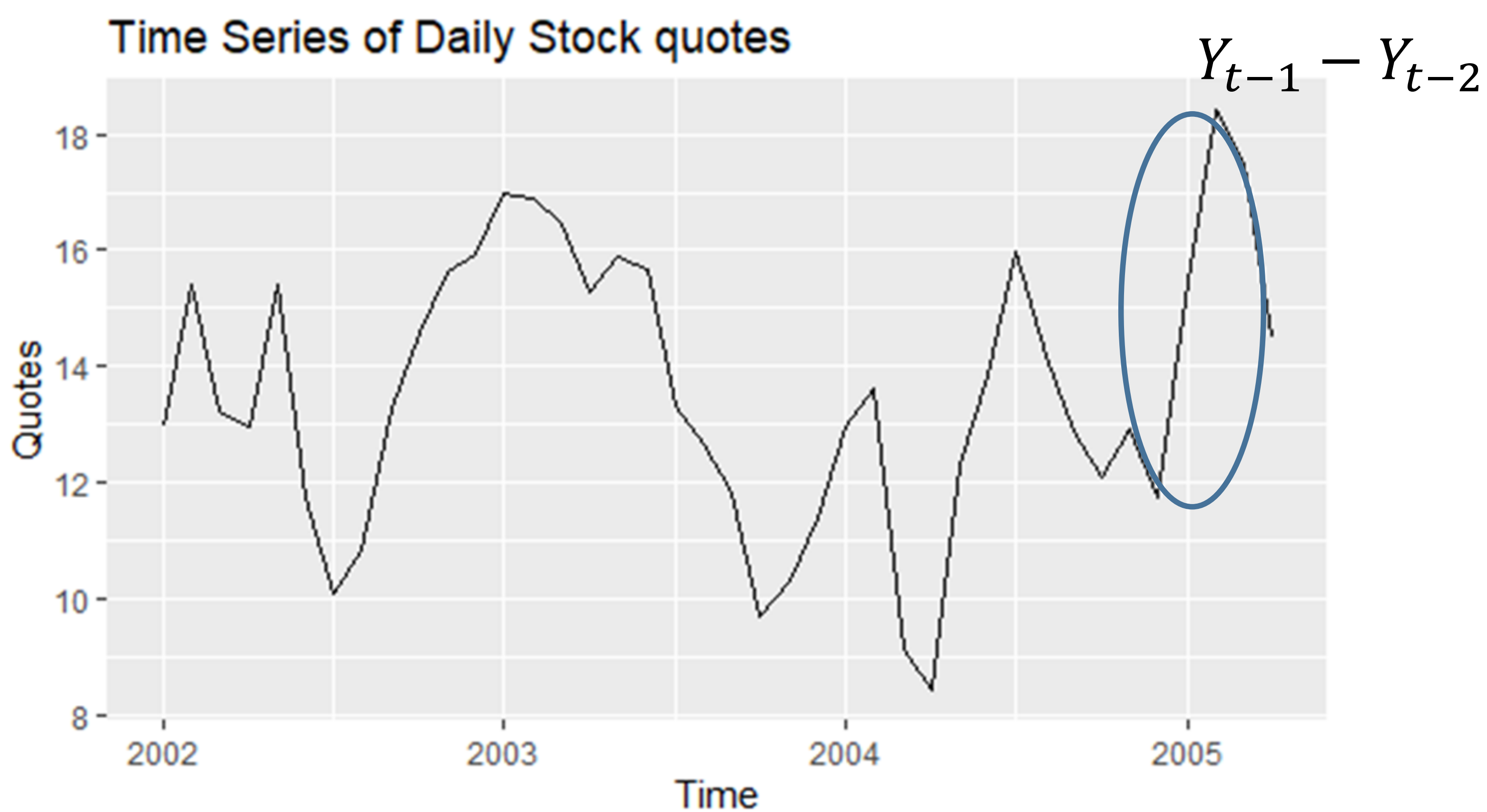
- General Model with Stochastic Trend:

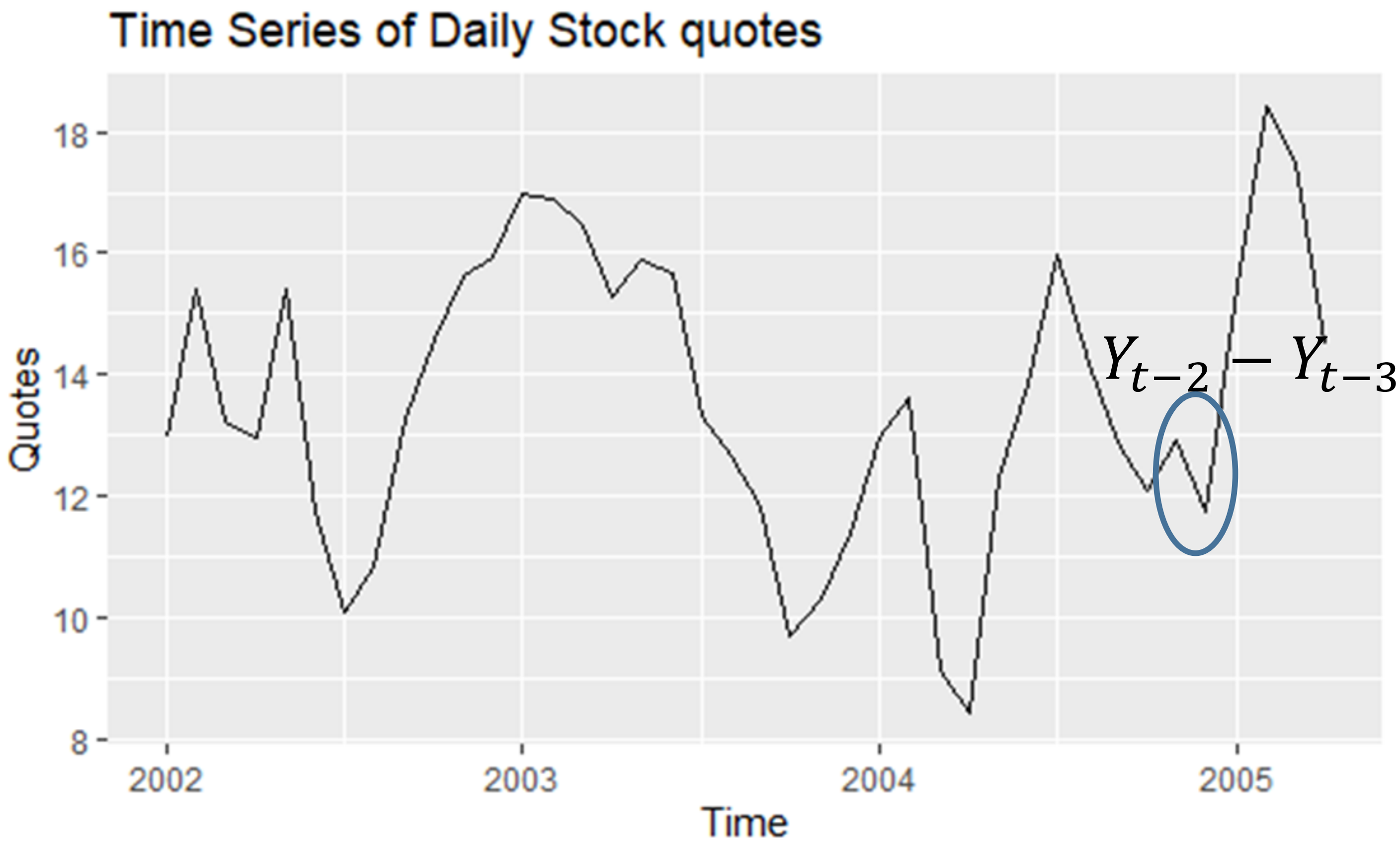
$$Y_t - Y_{t-1} = \varepsilon_t$$

Patterns may exist in the differences!

- Therefore, if a random walk exists, ***need*** to take difference of series

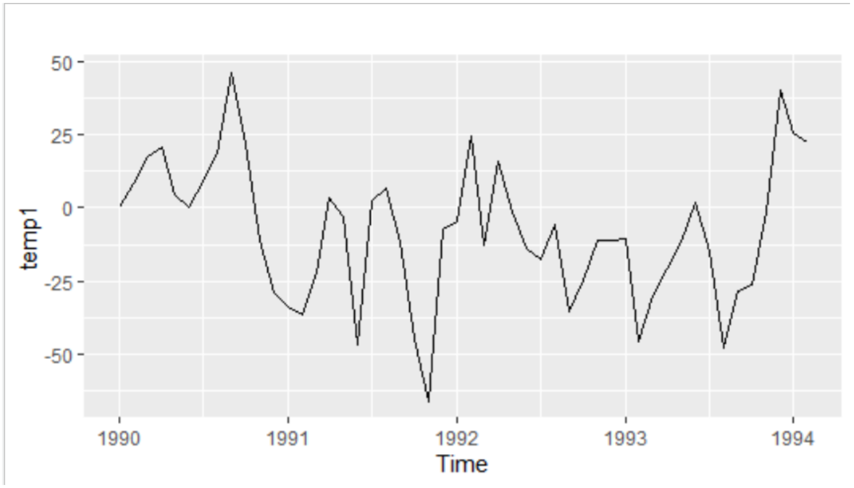

Time Series of Daily Stock quotes
$Y_t - Y_{t-1}$
Quotes
18
16
14
12
10
8
2002
2003
2004
2005
Time

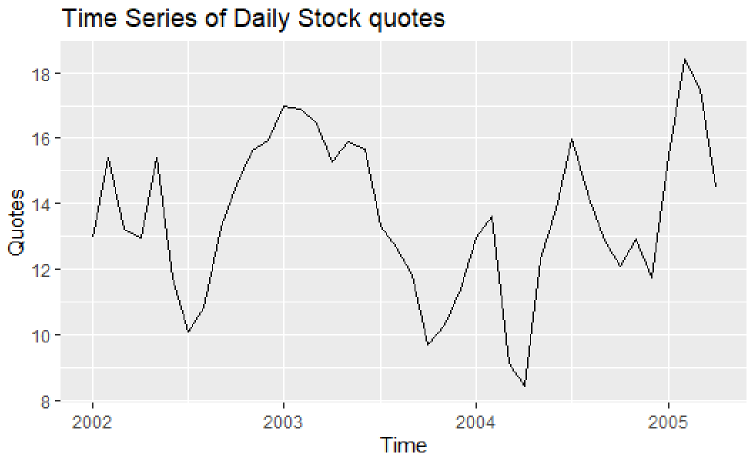

Time Series of Daily Stock quotes
$Y_{t-1} - Y_{t-2}$
Quotes
18
16
14
12
10
8
2002
2003
2004
2005
Time



Time Series of Daily Stock quotes
18
16
14
12
10
8
Quotes
2002
2003
2004
2005
Time
$Y_{t-2} - Y_{t-3}$

Example of two series (one with Random walk)





How do we know if we have a Random Walk or not?

UNIT ROOT TESTING

The Augmented Dickey-Fuller Unit Root Test

- This test provides a statistical test for detecting a random walk.
- The null hypothesis is that differencing is required (non-stationary data).
- The alternative hypothesis:
  1. Stationary mean about Zero (this is the test you will use if the series is centered about 0)
  2. Stationary Mean

The Dickey-Fuller Test

- Null Hypothesis:

Non-stationary! i.e…..Random Walk

- Alternative Hypothesis:

Stationary around 0 or mean (need to decide which test you are using based on the series)

Augmented Dickey-Fuller (ADF) Test

- It is called a "Unit Root Test" because it looks to see if the equation with the differenced series has a unit root (ϕ = 1….null hypothesis)

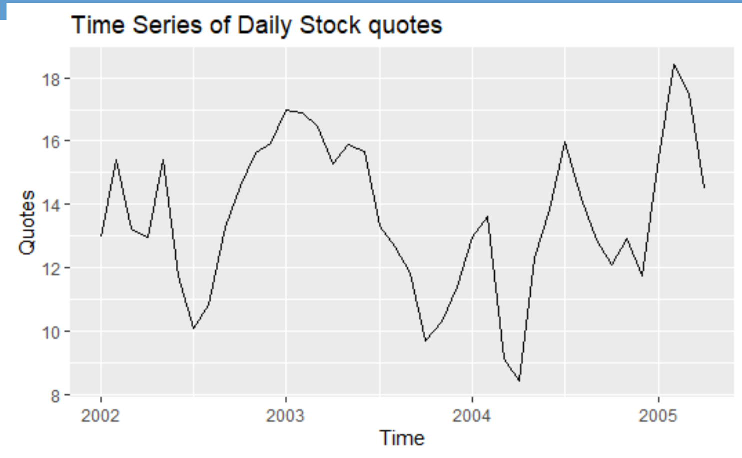
$$Y_t = \phi Y_{t-1} + e_t$$

- Unit roots can exist models with more than one lag of Y.
- Lag 0 tests are equivalent to what we have previously seen.
- Lag 1 tests consider models with differenced series of Y and first lag of differenced series.
- Lag 2 tests consider models with differenced series of Y and first and second lag of differenced series.
- When testing for stationarity, you should go to at least a lag 2 test (need to look at ALL of them…lag 0, lag 1 and lag2)

Augmented Dickey-Fuller Testing – R

```
# Augmented Dickey-Fuller Testing #
aTSA::adf.test(Quotes.ts)

### can also use ndiffs(Quotes.ts)…but be careful with this!!
```



Augmented Dickey-Fuller Test
alternative: stationary

ZERO MEAN

Type 1: no drift no trend

	lag	ADF	p.value
[1,]	0	-0.3061	0.550
[2,]	1	-0.5980	0.458
[3,]	2	-0.0632	0.620
[4,]	3	-0.0950	0.611

SINGLE MEAN

Type 2: with drift no trend

	lag	ADF	p.value
[1,]	0	-2.66	0.0939
[2,]	1	-3.42	0.0192
[3,]	2	-2.45	0.1608
[4,]	3	-2.36	0.1943

Type 3: with drift and trend

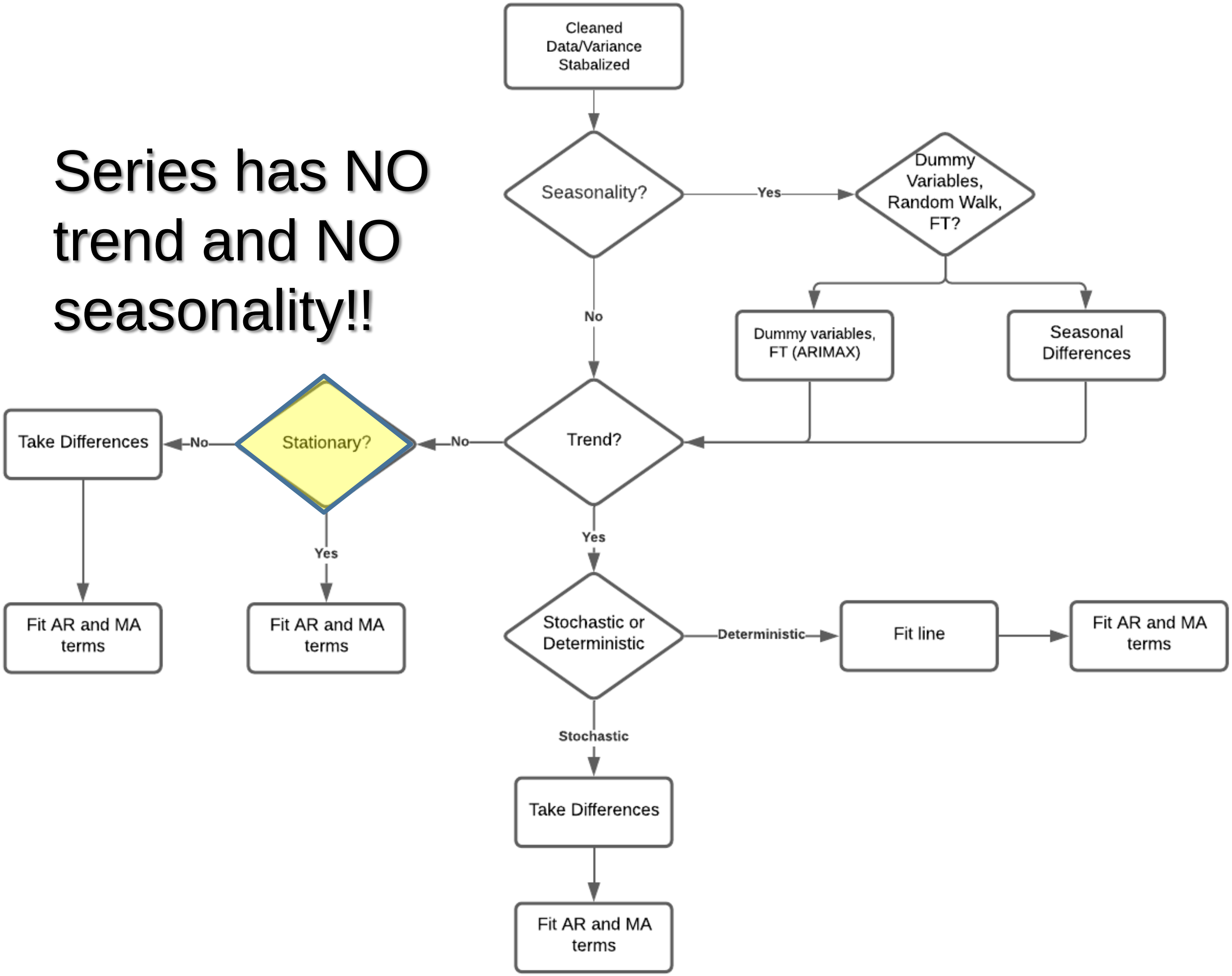
	lag	ADF	p.value
[1,]	0	-2.62	0.3212
[2,]	1	-3.36	0.0772
[3,]	2	-2.41	0.4012
[4,]	3	-2.29	0.4463

Note: in fact, p.value = 0.01 means p.value <= 0.01

ADF test:

1. First decide if you are doing the Zero Mean or Single Mean test (will talk about drift with trend in a few classes)
2. Decide how many lags you would like to look at (commonly done in industry is between 3-5 lags)
3. See if you reject ANY of these hypotheses!!
4. If you reject all hypotheses (i.e. feel confident that this is a stationary time series), then you are ready to start modeling AR and MA terms.
5. If you fail to reject at least one, you have a random walk and will take differences and start modeling AR and MA terms on the differences.

Series has NO trend and NO seasonality!!



Over-differencing

- When you difference and you don't need to difference, or you take too many differences, you will create the problem of **over-differencing**.
- This introduces more dependence on error terms in your model (creation of moving average terms that don't really exist).