ARMA

ARMA

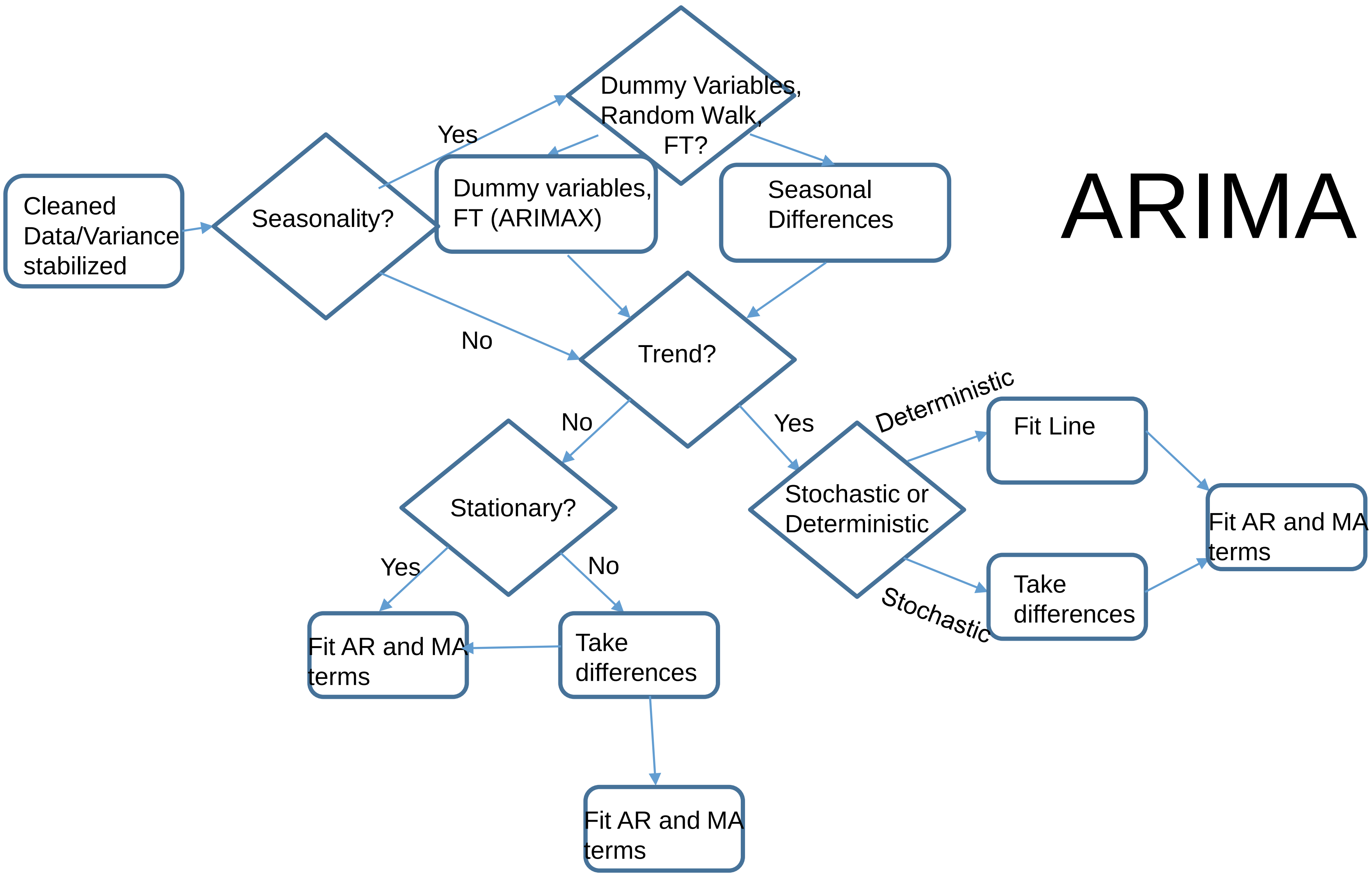
- ARMA stands for AutoRegressive Moving Averages
- ARMA models are based upon statistical methods (will assume a distribution!!)
- When creating ARMA models, it can be a circular process (when changing something later in a model might make you reevaluate what you did earlier)
- Best model will be found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

ARMA

SIGNAL:

- We will need to take care of ***functional form*** (obvious patterns, for example trend and/or seasonality)
- Also need to take care of correlation structure (this will come later!! Need to take care of functional form first!)

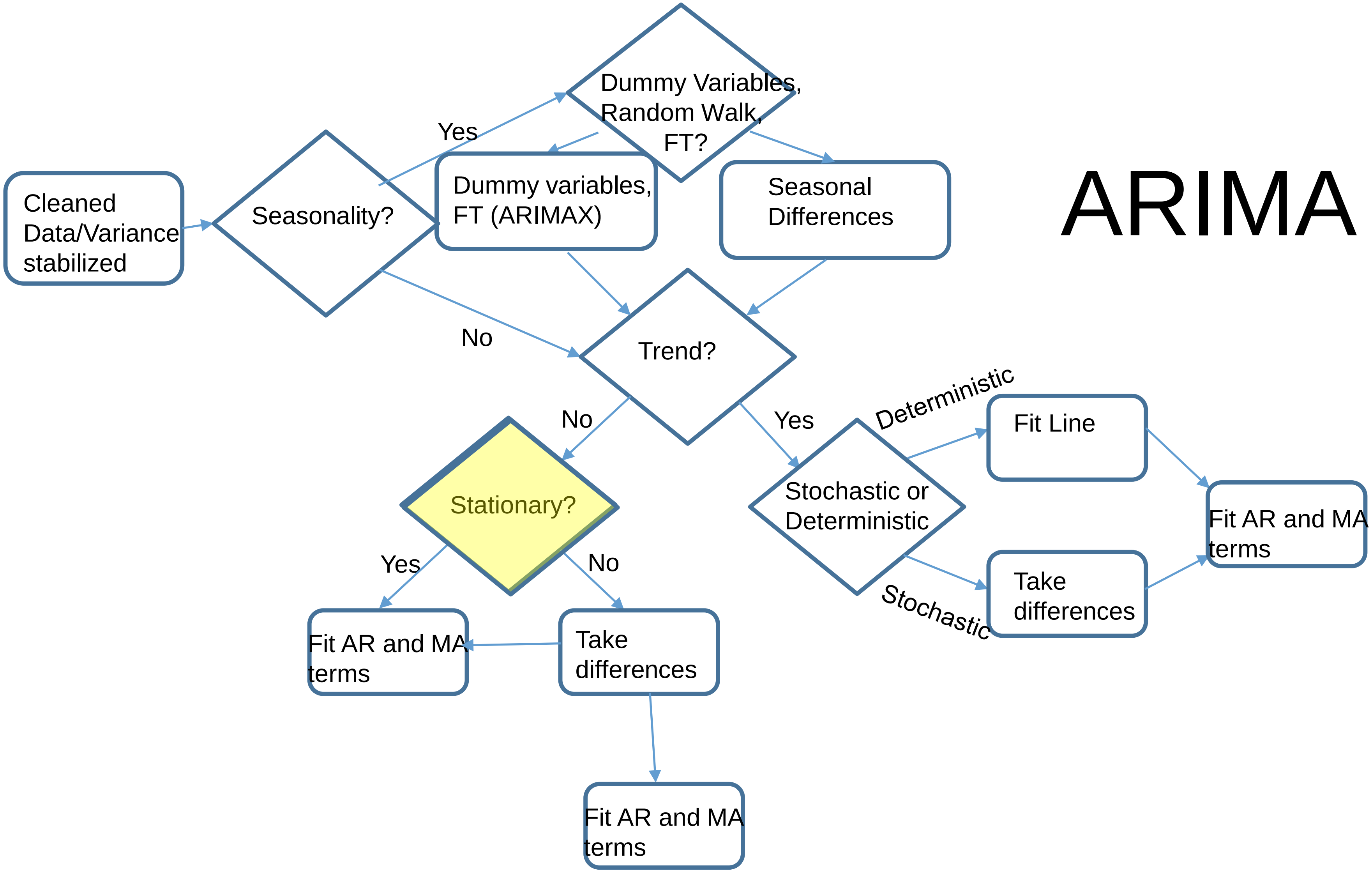
ARIMA



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

NO SEASON AND NO TREND (START SIMPLE….)

ARIMA



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

Stationarity

- To create an ARIMA model, we ***must*** have stationarity
- We will be using the idea of “weak stationarity” for modeling
  - No predictable pattern in the long-run and converges to a constant mean
  - A series with NO trend and NO seasonality will either be stationary or a random walk
  - Need to be able to identify random walks!

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:

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

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

Stochastic Trend: Differencing

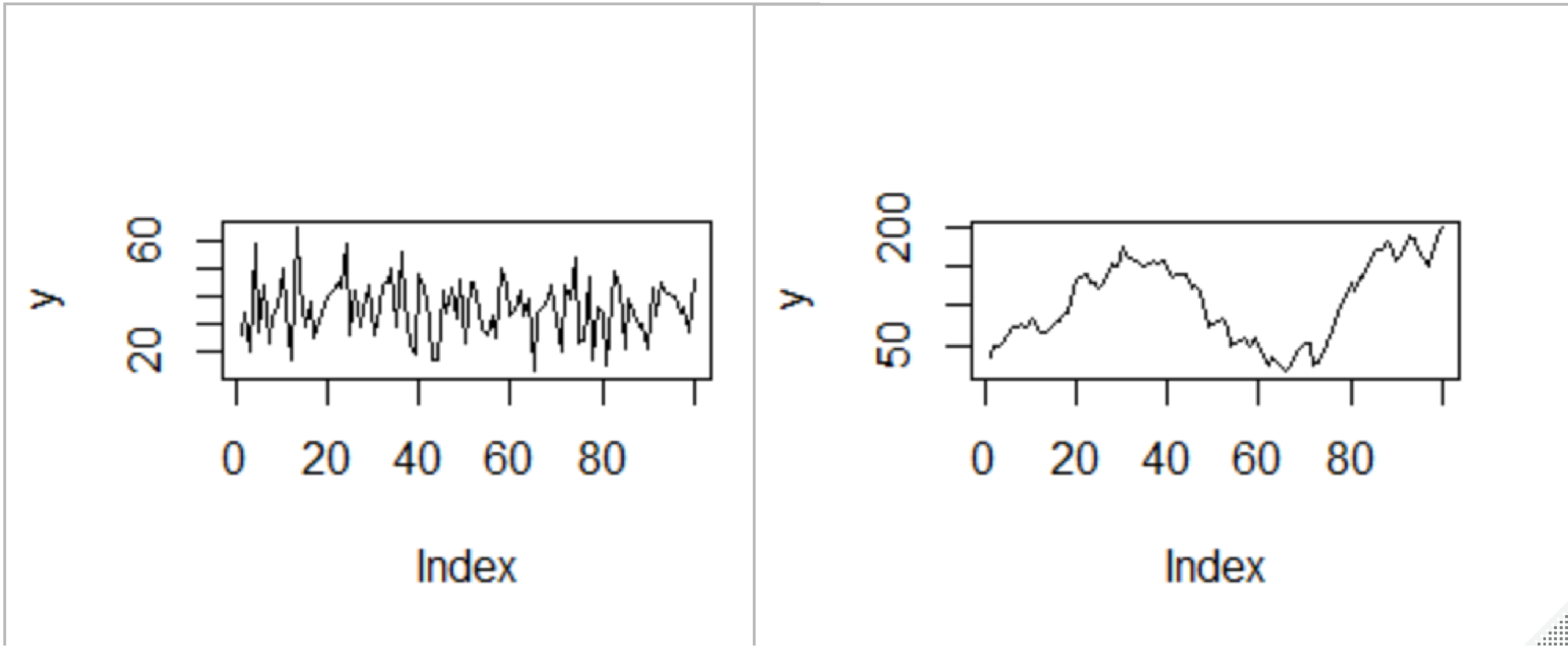
- General Model with Stochastic Trend:

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

Patterns may exist in the differences!

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

Example of two series (one with Random walk)



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

UNIT ROOT TESTING

The Dickey-Fuller Unit Root Test

- This test provides a statistical test for ***first*** differencing.
- The null hypothesis is that first differencing is required (non-stationary data).
- The alternative hypothesis:
  1. Zero Mean
  2. Single Mean

The Dickey-Fuller Test – Zero Mean

- Model:

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

- Null Hypothesis:

$$H_0\colon \phi = 1$$

- Alternative Hypothesis:

$$H_a\colon |\phi| < 1$$

The Dickey-Fuller Test – Zero Mean

- Model:

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

- Null Hypothesis:

$$H_0: \phi = 1$$

Non-stationary! i.e.….. Random Walk

- Alternative Hypothesis:

$$H_a: |\phi| < 1$$

Stationary around 0!

The Dickey-Fuller Test – Single Mean

- Model:

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

- Null Hypothesis:

$$H_0: \phi = 1$$

- Alternative Hypothesis:

$$H_a: |\phi| < 1$$

The Dickey-Fuller Test – Single Mean

- Model:

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

- Null Hypothesis:

$$H_0: \phi = 1$$

Non-stationary! i.e.….. Random Walk

- Alternative Hypothesis:

$$H_a: |\phi| < 1$$

Stationary around μ!

Augmented Dickey-Fuller (ADF) Test

- Unit roots are not limited to only random walk models with one lag of Y.
- Unit roots can exist models with more than one lag of Y.
- Higher order models are tested with the ADF tests.
- Lag 0 tests are equivalent to what we have previously seen.
- Lag 1 tests consider models with 2 lags of Y.
- Lag 2 tests consider models with 3 lags of Y.

Augmented Dickey-Fuller (ADF) Test

- Characteristic polynomial of an AR(p) model:

$$\phi(x) = 1 - \phi_1 x - \phi_2 x^2 - \cdots - \phi_p x^p$$

- Null Hypothesis:

$$H_0\text{: polynomial has root} = 1$$

- Alternative Hypothesis:

$$H_a\text{: polynomial is for stationary process}$$

Augmented Dickey-Fuller (ADF) Test

- The Rho test is the regression coefficient-based test statistic.
  - Superior power properties for Dickey-Fuller Test (ADF Lag 0 tests).
- The Tau test is the studentized test.
  - Superior power properties for all lags but ADF Lag 0 tests.
- The F test is the regression F test for the full model and the null hypothesis restricted reduced model, except that the distribution is not the usual F distribution used in ordinary regression.
  - Poorest power properties – seldom recommended.

Augmented Dickey-Fuller Testing – SAS

```
proc arima data=Time.fpp_insurance plot=all;
        identify var=quotes nlag=10 stationarity=(adf=2);
        run;
quit;
```

Augmented Dickey-Fuller Testing – R

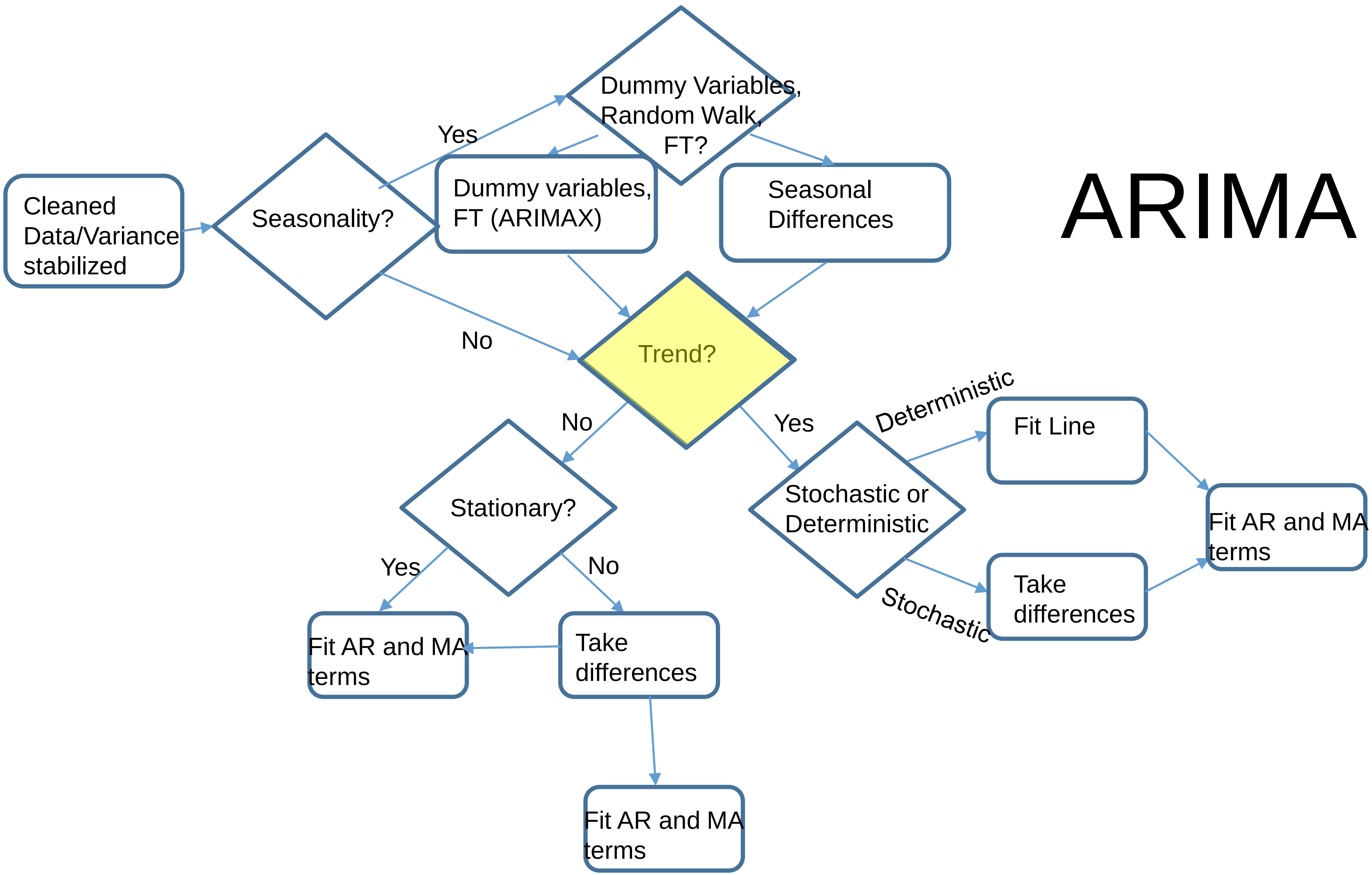
```
# Augmented Dickey-Fuller Testing #
adf.test(Quotes.ts, alternative = "stationary", k = 0)

### can also use ndiffs(Quotes.ts)
```

Augmented Dickey-Fuller Test

data: Quotes.ts
Dickey-Fuller = -2.6229, Lag order = 0, p-value = 0.329
alternative hypothesis: stationary

TRENDING DATA

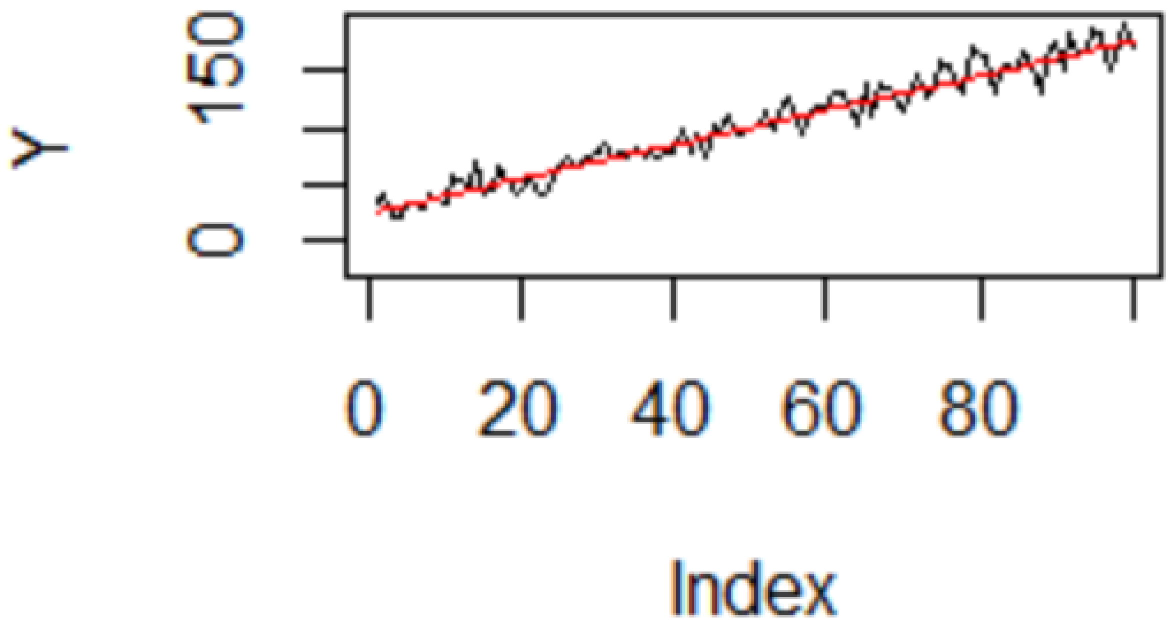
ARIMA


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

If you see a visible trend

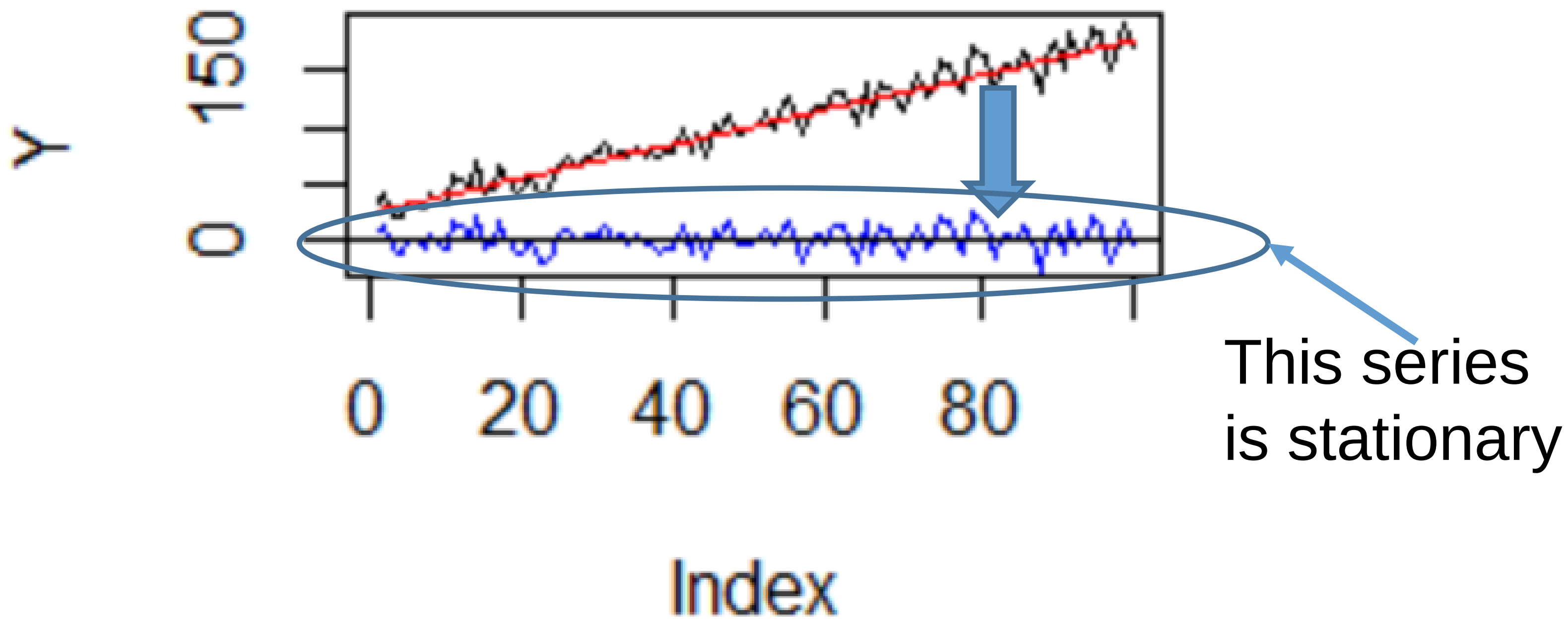
- If there is a trend, the current series is NOT stationary.
- Trending series are not stationary because they do not hover around a mean.
- One of two things can be happening:
  1. The series is stationary ABOUT THE LINE
  2. The series is a Random walk with drift

The series is stationary ABOUT THE LINE



Take away the trend and it is stationary!

Need to fit the trend line (residuals are stationary)



Deterministic Trends

- A deterministic trend is what we have done in regression:

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

- Where *t* is time
- Can also fit quadratic, exponential or any other form of time

Common Trend Models

- We are not limited to only having a linear trend:
  - Quadratic Trend:

$$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon_t$$

  - Logarithmic Trend:

$$Y_t = \beta_0 + \beta_1 \log(t) + \varepsilon_t$$

  - Exponential Trend:

$$Y_t = \exp(\beta_0 + \beta_1 t) + \varepsilon_t \rightarrow \log(Y_t) = \beta_0 + \beta_1 t$$

RANDOM WALK WITH DRIFT

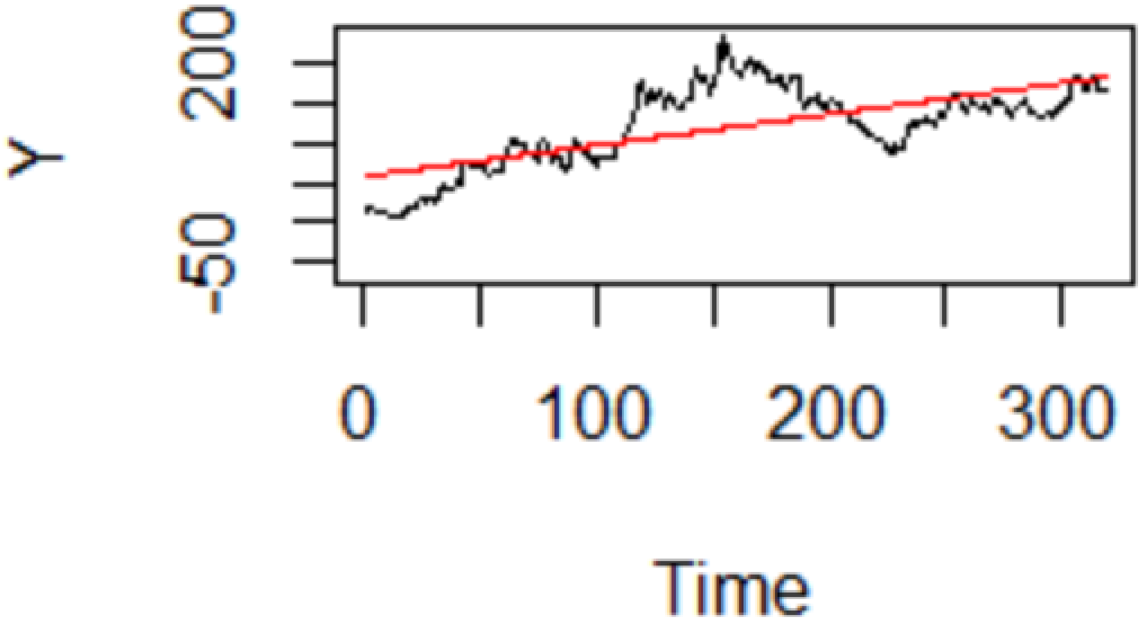
Random Walk with Drift Model

Random Walk with Drift

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

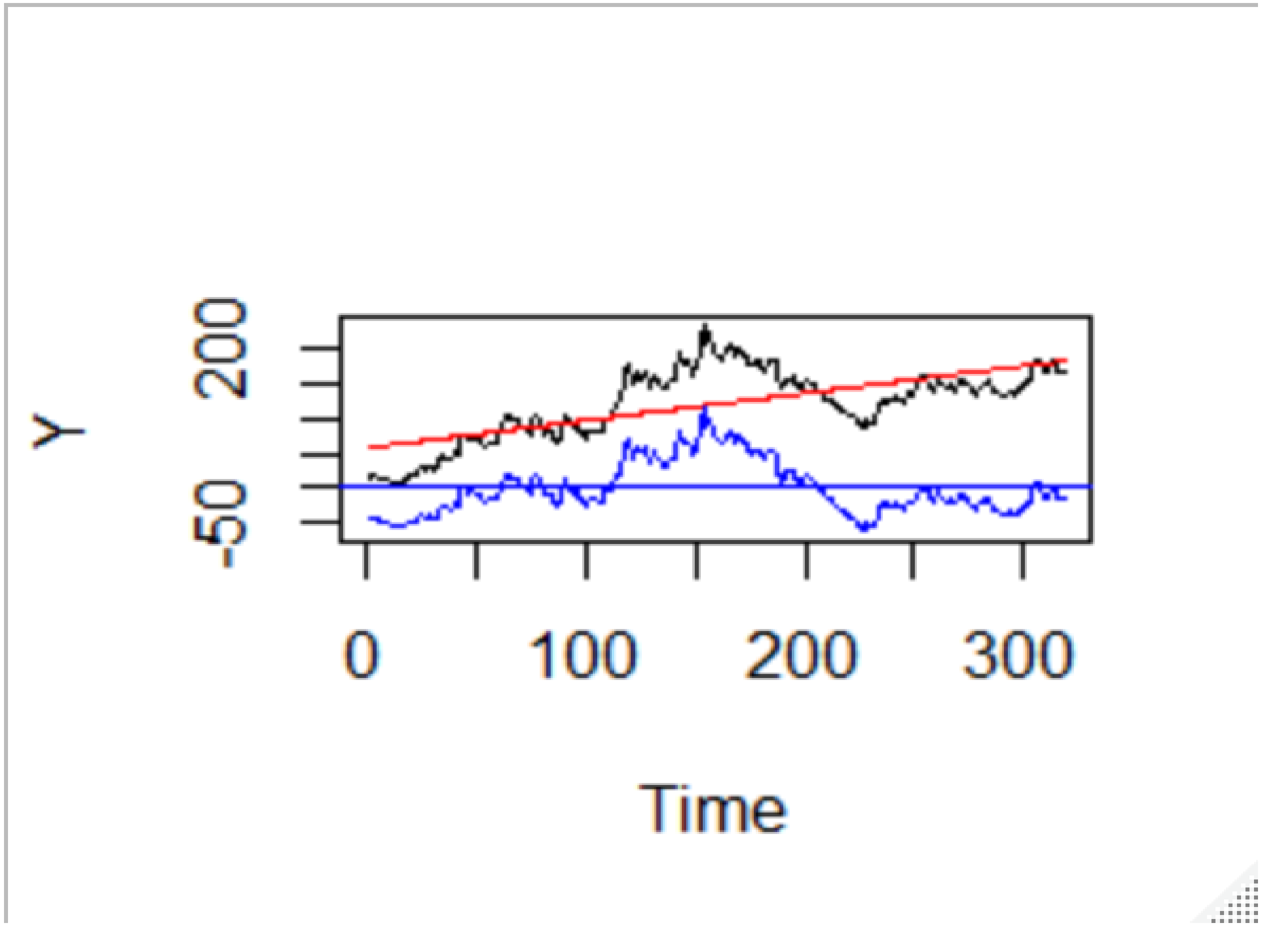
This controls the “drift” or the trend (if this is positive, it will “drift” upward; if it is negative, it will “drift downward)

Random Walk with Drift



Even if you remove trend line, the resulting residuals are NOT stationary!

Random Walk with drift is NOT stationary if you remove trend line!! Will need to take differences.



HOW CAN WE TELL?

The Dickey-Fuller Test – Trend

- Model:

$$Y_t - \beta_0 - \beta_1 t = \phi\big(Y_{t-1} - \beta_0 - \beta_1(t-1)\big) + e_t$$

- Null Hypothesis:

$$H_0: \phi = 1$$

- Alternative Hypothesis:

$$H_a: |\phi| < 1$$

The Dickey-Fuller Test – Trend

- Model:

$$Y_t - \beta_0 - \beta_1 t = \phi\big(Y_{t-1} - \beta_0 - \beta_1(t-1)\big) + e_t$$

- Null Hypothesis:

$$H_0: \phi = 1$$

Non-stationary! i.e…..
Random Walk

- Alternative Hypothesis:

$$H_a: |\phi| < 1$$

Stationary around trend!

The Dickey-Fuller Test – Trend

- Model:

$$Y_t - \beta_0 - \beta_1 t = \phi\big(Y_{t-1} - \beta_0 - \beta_1(t-1)\big) + e_t$$

- Null Hypothesis:

$$H_0: \phi = 1 \longleftarrow \text{Non-stationary!}$$

- Alternative Hypothesis:

$$H_a: |\phi| < 1 \longleftarrow$$

Deterministic trend,
NOT Stochastic trend

When an obvious trend exists

- The series is ***NOT*** stationary.
- Need to determine if it is a deterministic trend OR a stochastic trend (random walk with drift)
  - If it is a deterministic trend, fit a regression line and then use residuals for rest of analysis (part of ARIMAX)
  - If it is a random walk (stochastic), take first difference
- Examples for each situation follows…

SAS Code

```
proc arima data=Time.usairlines plot=all;
        identify var=passengers stationarity=(adf=2) ;
run;
quit;
```

R Code

```
adf.test(Passenger,alternative = "stationary", k = 0)
arima.trend=Arima(Passenger,xreg=x,order=c(0,0,0))
```

RANDOM WALK WITH DRIFT

SAS Code

```
proc arima data=Time.Ebay9899 plot=all;
	identify var=DailyHigh nlag=10 stationarity=(adf=2);
	identify var=DailyHigh(1) nlag=10 stationarity=(adf=2);
run;
quit;
```

SAS Output

```
proc arima data=Time.Ebay9899 plot=all;
        identify var=DailyHigh nlag=10 stationarity=(adf=2);
        identify var=DailyHigh(1) nlag=10
stationarity=(adf=2);
run;
quit;
```

R code

```
Daily.High <- ts(Ebay$DailyHigh)
Daily.High<-Daily.High %>% na_interpolation(option = "spline")
adf.test(Daily.High,alternative = 'stationary',k=0)
rw.drift=Arima(Daily.High,order=c(0,1,0))
summary(rw.drift)
```

```
Series: Daily.High
ARIMA(0,1,0)

sigma^2 estimated as 47.76:  log likelihood=-1062.6
AIC=2127.21   AICc=2127.22   BIC=2130.96

Training set error measures:
                    ME     RMSE      MAE       MPE     MAPE      MASE
Training set 0.4779789 6.900166 4.895049 0.4782673 4.541625 0.9968669
                  ACF1
Training set 0.2026237
```

Over-differencing

- When you difference in the presence of deterministic trends, or you take too many differences in a stochastic trend 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).