INTRO

- FORECASTING → PREDICT THE FUTURE USING DATA IN HAND.

BASICALLY, YOU CAN FORECAST EVERYTHING.

WHY? → IMPORTANT IN THE BUSINESS DECISION-MAKING.

↓

IT AFFECTS DECISIONS WE MAKE TODAY.

MOST OF THE TIME IS WRONG, THEREFORE, IT IS ACCURATE MORE IN THE SHORT TERM THAN IN THE LONG ONE.

→ MORE COMPLETE THE DATA IS, THE BETTER A DECISION WILL BE.

→ IF EFFECTIVE, IT WILL BRING TO

1. COSTS REDUCTION
2. CUSTOMER SATISFACTION

ISSUES

- FIRM POLITICS
- OVERCONFIDENCE
- WISHFUL THINKING
- SUCCESS // FAILURE ATTRIBUTION
- GAMBLER'S FALLACY
- DATA PRESENTATION
- CONSERVATISM

- COSTS ☆
- TIME ☆
- ACCURACY ☆

INFLUENCING FACTORS

EXTERNAL

- ECO. CONDITIONS
- REGULATIONS
- COMPETITION
- CONSUMER BEHAVIOR

INTERNAL

- PRICE
- PROMOTION
- PRODUCT
- QUALITY

PROCESS

WHAT BRINGS TO A GOOD FORECAST.

1. TIMELY
2. RELIABLE
3. ACCURATE
4. MEANINGFUL UNITS
5. WRITTEN
6. UNDESTANDABLE

* MAX WE CAN STRETCH IS NORMALLY A YEAR.

!

- BEFORE U START FORECASTING, YOU SHOULD ANSWER WHY FIRST .....

1) WHY YOU NEED A FORECAST ? [PURPOSE]

2) COLLECT HISTORICAL DATA [PATTERNS]

    - CAUSAL FORECASTING
    - UNIVARIATE FORECASTING

3) PLOT DATA

4) SELECT A MODEL

5) DEVELOP FORECAST FOR CHOSEN PERIOD

6) IS ACCUDATE ?

NO

YES, MOVE ONE

7) PLAN HORIZON

8) ADJUST AND PERFECT THE MODEL.

DATA

STATIONARY

STABLE OVERTIME

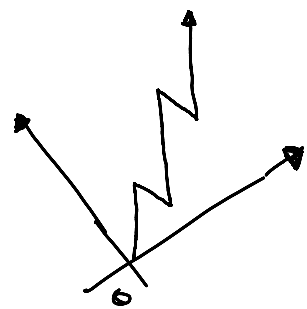
NON STATIONARY

UPWARD AND DOWNWARD TRENDS

SEASONAL

SHOWS PATTERNS OVER TIME

- ① DETERMINISTIC
  - ↳ EASILY TO PREDICT (SUNRISE SUNSET)
- ② RANDOM → UNPREDICTABLE NO PATTERN ⚠



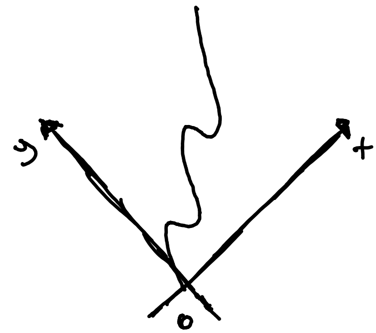
PROCEDURE

1. CHOICE → MSE, → MAD
2. ACCURACY → MAPE
3. TRACKING SIGNAL

COMPONENTS OF A FORECAST

- TREND → DANIEL TEST
- CYCLE → KRUSTAL - WALLIS TEST

↳ SEASONALITY OR RANDOMNESS



ERRORS OR DEVIATIONS

THE GOAL IS TO MINIMAZE IT.

↳ POSITIVE ERRORS ARE MORE THAN NEGATIVE ONE.

TO BOOST THE ACCURANCY, SPLIT IN 2 THE TIME SERIES.

→ [SAMPLE PERIOD]

→ [BEYOND SAMPLE PERIOD]

- IF THERE'S A CONTRADICTION, $\boxed{\text{MSE}}$ HAS THE PRIORITY UPON $\boxed{\text{MAD}}$.

MAPE = MEAN ABSOLUT PERCENT ERROR

$$\frac{\sum \left| \frac{e_{j+T}}{y_{T+i}} \right|}{H}$$

IN THIS CASE I GET

$$\frac{\sum \frac{\text{ERROR OF REAL VALUÉ}}{\text{REAL VALUE}}}{n}$$

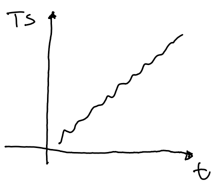
INTERPRETATION

$$\begin{cases} \text{MAPE} \leq 1\% & \text{VERY GOOD} \\ 1\% < \text{MAPE} = 3\% & \text{GOOD} \\ 3\% < \text{MAPE} = 5\% & \text{REGULAR} \\ 5\% < \text{MAPE} & \text{LOW} \end{cases}$$

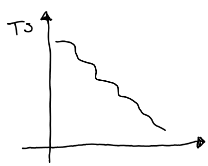
TRACKING SIGNAL

CALCULATES THE ERROR'S BEHAVIOUR.

$$T.S. = \frac{\sum (R.V - P.V)}{MAD}$$



POSITIVELY BIASED
(TREND)

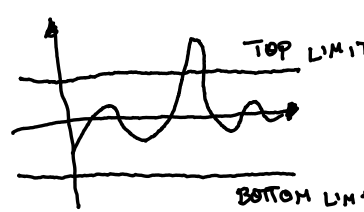


NEGATIVELY BIASED
(TREND)

- KEEP IN MIND, IN AN EVALUATION, DON'T FORGET TO TRACK THE TRENDS ⚠

- THE MAD HAS TO BE CALCULATED FOR EACH CUMULATIVE REAL VALUE OF THE FORECASTING SAMPLE.

- A FORECASTED ERROR HAS TO BE IN 2 LIMIT'S RANGE.



TREND

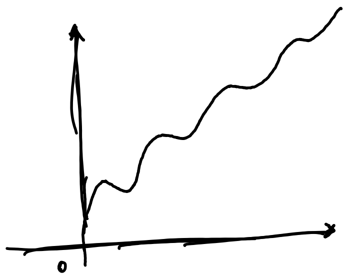
HOW A DATA SET BEHALF IN THE LONG-RUN.

- POPULATION GROWTH
- TECHNOLOGY
- INNOVATION
- CONSUMER PREFERENCES

- CYCLICAL FLUCTUATION
  - BUSINESS AND ECONOMY
    - FOR EXAMPLE, WHEN YOU LAUNG A PRODUCT.
  - (FACTORS)
    1. INFLACTION // DEFLACTION
    2. STOCK MARKET PRICES
    3. UNIQUE EVENTS

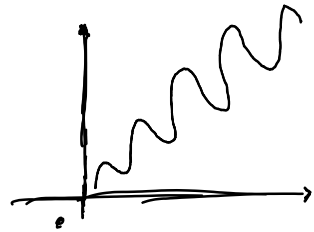
AGGREGATION

TREND + SEASONALITY



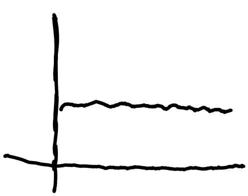
→ STANDARD DEVIATION STALLS

TREND × SEASONALITY



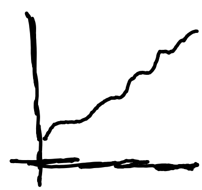
→ STANDARD DEVIATION INCREASES

TYPE 1



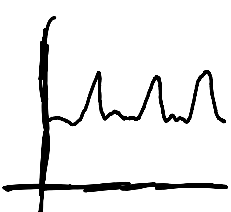
NO TREND
NO SEASON

TYPE 3



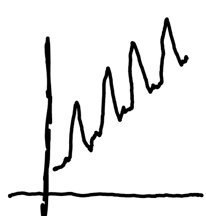
WITH TREND
NO SEASON

TYPE 2



NO TREND
WITH SEASON

TYPE 4



WITH TREND
WITH SEASON

THE DANIEL TEST → CHECK TRENDS

1. HYPOTHESIS → H_0 : NO, TREND
   H_A : YES, TREND

2. STATISTICS → FIND DISTRIBUTION

3. DECISION → RH_0 → T ✓
   not RH_0 → T ✗

→ FORECAST AND REASON OF OF THOSE FOUR TYPES.

$$\boxed{\tau_s = 1 - \frac{6 \cdot \sum d_t^2}{T(T^2-1)}} \sim N\left(0, (T-1)^{-1}\right)$$

WE HAVE TO ADJUST IT

⚠

- IF WE HAVE A NEGATIVE VALUE, THERE'S A NEGATIVE TREND.
- SAME APPLIED TO THE POSITIVE.

$$Z = \sqrt{T-1} \cdot \tau_s$$

→ $N(0, 1)$

KRUSKAL - WALL'S TEST

→ CHECK SEASONALITY

- SAME PROCESS AS DANIEL TEST, BUT WITH A DIFFERENT FORMULA.

$$\boxed{H = \frac{12}{T(T+1)}\left[\sum \frac{R^2}{T}\right] - 3(T+1)}$$

- CONFRONT THE FORMULA WITH

CHI-SQUARED → $\chi^2_{\alpha}(s-1)$

→ # SEASONS

QUARTS, MONTHS, SEMESTERS

IF $H > \chi^2$ → H_0 IS REJECTED (NO SEASONS)

IF $H < \chi^2$ → H_0 IS ACCEPTED (YES SEASONS)

1^{ST} TYPE

① NAIVE FORECASTING

↳ I STICK WITH THE PREVIOUS VALUE IN ORDER TO FORECAST.

② SIMPLE MEAN

↳ APPLY THE ARITHMETHIC MEAN $\bar{X} = \frac{\sum x_i}{n}$

③ MOVING AVERAGE

↳ CHOOSE THE [LENGTH] OF # VALUES

SO, YOU MOVE TOWARD BY REMOVING THE LAST NUMBER OF THE ROSTER AND ADD THE FOLLOWING ONE.

$\boxed{K}$ → IF REALLY SMALL, USE NAIVE METHOD.

RESPONSIVE ↑ ↓ STABLE

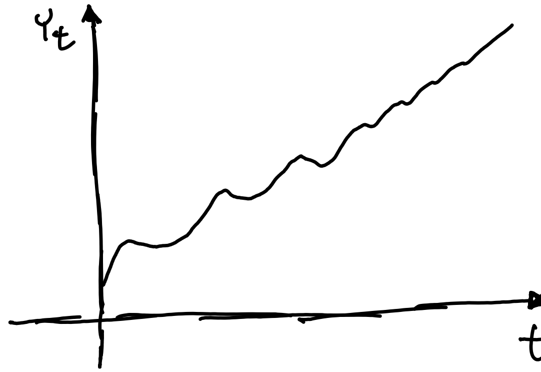
↳ IF EXTREMELLY LARGE, USE SIMPLE MEAN.

→ IF YOU HAVE DIFFERENT VALUES, CALCULATE $\boxed{MSE}$ AND CHOOSE THE ONE WITH THE SMALLER MSE.

④ SIMPLE EXPONENTIAL SMOOTHING

↳ LOOK UP THE SLIDES FOR A BETTER UNDERSTANDING ON HOW TO USE SUCH METHOD.

3RD TYPE



1. LINEAR TREND

↳ $\beta_0 + \beta_1 \cdot t = \hat{y}$

LOCALLY LINEAR →

2. DOUBLE MOVING AVERAGE
3. HOLT EXPONENTIAL SMOOTHING

KEEP IN MIND, DANIEL TEST HAS FLAWS. THEREFORE, ONE SHOULD BETTER DRAW THE GRAPH AND SEE WHAT IS GOING ON, OTHERWISE THE FORECAST COULD BE WRONG.

TO FORECAST BEYOND SAMPLE PERIOD

↳ $\beta_0 + \beta_1 \cdot (T + m)$

STOCHASTIC PROCESS

- **WHITE NOISE**

↳ SIMPLEST AND BASIC METHOD

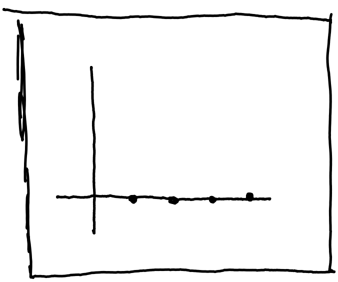
- $E(U_t) = \varnothing$ → ERRORS MOVE AROUND $\varnothing$

- VAR = CONSTANT

- COV = $\varnothing$ → THERE ARE NO PATTERNS. THUS NO CORRELATION BETWEEN THE TWO VALUES.

[AUTO - CORRELATION FUNCTION]

ACF



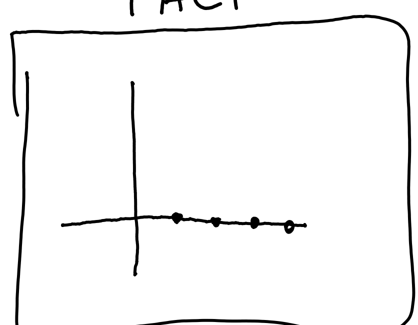
ALL VALUES ARE $\varnothing$ DUE TO VARIANCE

$$\rho_J = \frac{VAR_J}{VAR_\varnothing}$$

[PARTIAL]

PACF

IT IS LIKE THE $\boxed{\text{ACF}}$



- **RANDOM WALK**

↳ ADDING A CONSTANT TO WHITE NOISE

↳ CALLED

LAG-OPERATOR

$L(x_t)$

- WHAT IT DOES,

IT THROWS IN THE

PAST OUR TIME-SERIES

A RANDOM WALK IS A NON-STATIONARY MODEL ⚠

$$E(RW) = E(\delta + w_t + y_{t-1})$$

$$\vdots$$

$$= E(t) \cdot \delta + E(y_0)$$

$$\vdots$$

$$= E(y_0) + t\delta$$

IT DEPENDS

ON TIME UNLESS SOME VALUES ARE $\emptyset$

MOVING AVERAGE, ORDER 1

$$Y_t = \mu + w_1 - \theta_1 w_{t-1}$$

WHERE w_1 AND w_{t-1} ARE WHITE NOISES.

$\underline{MA}\ (1) \rightarrow$ HOW MANY LAG VALUES APPLIED.

$\hookrightarrow$ MODELRY ERRORS

- LET'S APPLY SOME MATH

$$Y_t = \mu + w_t - \theta_1 w_{t-1}$$

$$= \mu + w_t - \theta L w_{t-1}$$

$$= \mu + \underline{(1 - \theta_1 L)}\, w_t$$

* LOOK AT SLIDES TOPIC 5 TO CONSIDER CHARTS ⚠

ARIMA

INVERTIBILITY

AUTO - REGRESSIVE → LAGS TIME SERIES (y_t)

MOVING AVERAGE → LAGS ERRORS (w_t)

ARIMA (_ ; _ ; _)

WHITE NOISE → ARIMA (0 ; 0 ; 0)

RANDOM WALK → ARIMA (0 ; 1 ; 0)

SARIMA EXERCISE

$$\overset{B}{\underset{\text{DIFF R}}{(1-L)}}\ \overset{C \ \text{AR SEA}}{\underbrace{(1-0.5L^4-0.8L^8)}_{\text{SEASONAL}}}\ \overset{D}{\underbrace{(1-L^4)^2}_{\text{DIFFER SEASON}}}\ \overset{A}{\underbrace{(1-0.73L)}_{\text{REGULAR}}}\ y_t = 3.8 + \overset{\text{MA (1)} \ E}{\underbrace{(1-0.3L-0.7L^2)}_{\text{REGULAR}}}\ \overset{\text{MA (1)} \ F}{\underbrace{(1-0.85L^4)}_{\text{SEASONAL}}}\ w_t$$

- STATIONARY (AR) $\begin{cases} R \\ S \end{cases}$
- INVERTIBLE (MA) $\begin{cases} R \\ S \end{cases}$

$$\text{SARIMA}\ \left(\overset{A}{\underline{1}} ; \overset{B}{\underline{1}} ; \overset{E}{\underline{2}}\right)\left(\overset{C}{\underline{2}} ; \overset{D}{\underline{2}} ; \overset{F}{\underline{1}}\right)$$

- ONCE DONE, THE SECOND STAGE IS TO CHECK STATIONARITY EITHER FOR TIME SERIES AND ERRORS.

$$\text{RULES} \begin{cases} \phi_1 + \phi_2 < 1 \\ \phi_1 - \phi_2 < 1 \\ |\phi_1| < 1 \end{cases}$$

IF JUST ONE IS BUSTED.

IT MIGHT BE NOT STATIONARY OR INVERTIBLE

BOX - JENNINS MET

1 IDENTIFICATION

- ANALYZE STATIONARITY → IN MEAN / → IN VARIANCE

2 ESTIMATION

- ORDINARY LEAST SQUARES (OLS)
- MAXIMUM LIKELIHOOD (ML)

3 DIAGNOSIS

- CHECK ADEQUACY →
  1. APPLY TOO MANY DIFFERS.
  2. APPLY TOO FEW DIFFERS.
  3. FORGOT A MODEL
  4. TOO MANY MODELS OR THE ORDER IS TO HIGH

4 APPLICATION

SIGNIFICANCY LEVELS

$$\left\{\begin{array}{lcl} *** & \rightarrow & 1\% \\ ** & \rightarrow & 5\% \\ * & \rightarrow & 10\% \end{array}\right.$$

THOSE STARS ARE USED TO REMOVE NON SIGNIFICANT DATAS FROM YOUR MODEL.

$$\left\{\begin{array}{ll} \text{IF} & \text{P-VALUE} < \text{SIGN}_{\text{LEVEL}} \quad (\text{SIGNIFICANT}) \\ \text{IF} & \text{P-VALUE} > \text{SIGN}_{\text{LEVEL}} \quad (\text{NON SIGNIFICANT}) \end{array}\right.$$

- THUS, THE MOST NON-SIGNIFICANT DATA, HAS TO BE REMOVED.

WHILE PERFORMING THE DIAGNOSIS, A MODEL, IN ORDER TO BE ADEGUATE HAS TO HAVE 4 CHARACTERISTICS.

1. ALL COEFFICIENTS HAVE TO BE SIGNIFICANT; (LOOK AT CORRELOGRAMS)
2. STATIONARITY AND INVERTIBILITY CONDITIONS; (BOX BELOW)
3. RESIDUALS HAVE TO FOLLOW A WHITE NOISE; (P-VALUE > 0.05)
4. STABILITY; ✓

MODEL	p	q	STATIONARITY	INVERTIBILITY
ARMA (1;0)	1	0	$\lvert\phi_1\rvert < 1$	NO
ARMA (1;1)	1	1	$\lvert\phi_1\rvert < 1$	$\lvert\theta_1\rvert < 1$
ARMA (2;0)	2	0	$\lvert\phi_1\rvert < 1$ $\phi_1 + \phi_2 < 1$ $\phi_2 - \phi_1 < 1$	NO
ARMA (0;1)	0	1	NO	$\lvert\theta_1\rvert < 1$
ARMA (0;2)	0	2	NO	$\lvert\theta_1\rvert < 1$ $\theta_2 + \theta_1 < 1$ $\theta_2 - \theta_1 < 1$

DURING COMPARISONS BETWEEN THE DATA SET THE ORDER IS GIVEN BY THE MISSING VALUE WITHIN THE RANGE.

OCTOBER - NOVEMBER = ORDER 1

JANUARY - MARCH = ORDER 2

JANUARY - JUNE = ORDER 5

GRETL

- PLOT A GRAPH TO CHECK WHETHER THE DATA-SET IS STATIONARY IN ITS MEAN OR NOT.

- IF NOT, PERHAPS DIFFERENCES OR LOGS HAVE TO BE APPLIED IN ORDER TO MAKE IT STATIONARY.

KEEP IN MIND, ONE WANTS TO APPLY DIFFERENCES TO A SERIES TO FLATTEN IT, WHEREAS LOGS ARE USED TO WITHER VARIANCE.

TYPE 1 NO TREND NO SEAS

- SIMPLE MEAN
- NAIVE
- MOVING AVERAGE
- EXPONENTIAL SMOOTHING

TYPE 2 NO TREND

- NAIVE
- MEAN

TYPE 3

- LINEAR TREND
- D. MOVING AVERAGE
- HOLT EXPONENTIAL

TYPE 4 BOTH

- DECOMPOSITION
- HOLT - WINTERS

EXERCISE 1

$\boxed{\text{DEVIATION}}$ = REAL VALUE − PREDICTED VALUE

TIME	METHOD A	ERROR		METHOD B	ERROR
1999	2150	-30		2010	110
2000	2345	-45		2278	22
2001	2800	-75		2650	75

REAL VALUES → 2120 ; 2300 ; 2725

(A) METHOD B PROVIDES THE BEST FORECASTS, BECAUSE THE OBJECTIVE IS TO MITIGATE THE DEVIATIONS. IN THIS CASE, $\boxed{B}$ HAS POSITIVE ERRORS.

(B) MAD → MEAN ABSOLUTE DEVIATION

$$\text{METHOD A} = \frac{\sum |RV - PV|}{n} = 50 \qquad \text{METHOD B} = 69$$

MES → MEAN SQUARED ERROR

$$\text{METHOD A} = \frac{\sum (RV - PV)^2}{N} = 2850 \qquad \text{METHOD B} = 6069.66$$

IN THE EVALUATION, METHOD A STRICTLY PROVIDES BETTER FORECASTS THAN METHOD B ⚠