QBUS6840 Lecture 02

Data Pattern, Graphing, Time Series Components, and Forecast Accuracy

Professor Junbin Gao

The University of Sydney Business School

Outlines

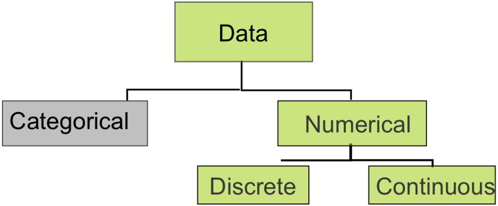
- Data and Data Pattern
- Graphing Data
- Components of Time Series
- Naïve Methods
- Moving Average
- Prediction Error and Measures
- Assessing and Choosing Models

Readings: Online textbook Chapter 2: `otexts.com/fpp2/graphics.html` and Chapter 3: `otexts.com/fpp2/simple-methods.html`

Data

- What are data? What data would be useful? Is it available? Are there any special events?
- Data carry information/knowledge: numeric, image, video, sound, text, any forms you can think of.
- Traditionally data collected to prove some 'hypotheses", to infer knowledge or causality
  - The Australian Safety Commission wants to measure the safety on Qantas flights after some recent near mid-air collisions. For the next 3 months they count the number of dangerous incidents involving Qantas flights.
  - A pharmaceutical company conducts a study on effectiveness of its new painkilling drug through trials.
  - A financial analyst wants to predict stock returns based on company accounting variables. A sample of companies is randomly obtained from the ASX.
- Today most data collected in passive ways thanks to digital technology.

Data Types and Categories



- Structured Data
- Unstructured Data
- Text data
- Graph Data
- Visual Data
- Audio Data
- ... ...

Levels of Measurement and Measurement Scales

- Nominal: Nominal scales are used for labeling variables, without any quantitative value. A good way to remember all of this is that ‘nominal’ sounds a lot like ‘name’ and nominal scales are kind of like ‘names’ or labels.
- Ordinal: With ordinal scales, it is the order of the values is what’s important and significant, but the differences between each one is not really known.
- Interval: Interval scales are numeric scales in which we know not only the order, but also the exact differences between the values.
- Ratio: they tell us about the order, they tell us the exact value between units, AND they also have an absolute zero which allows for a wide range of both descriptive and inferential statistics to be applied.

Levels of Measurement and Measurement Scales

	Nominal	Ordinal	Interval	Ratio
The "order" of values is known		X	X	X
Counts or Frequency	X	X	X	X
Mode	X	X	X	X
Median		X	X	X
Mean			X	X
Difference between each value			X	X
Add and Abstract			X	X
Multiplication and Division				X
True "Zero"				X

Graphing Data: Great Examples

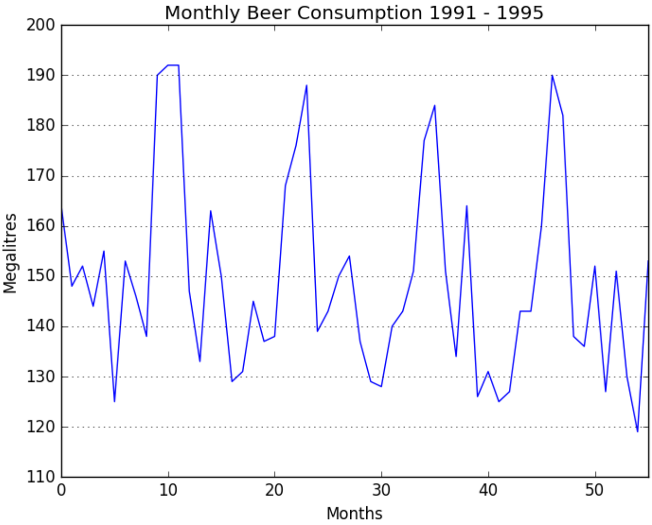
- Hans Rosling's video from youtube: 200 Countries, 200 Years, 4 Minutes
  http://www.youtube.com/watch?v=jbkSRLYSojo&feature=youtu.be
- Google Charts
  https://google-developers.appspot.com/chart/interactive/docs/gallery
- Single most important thing for learning about data patterns!
- Type of data $\Rightarrow$ type of graph
  Ideal graphs $\Rightarrow$ Ideal graphs convey both patterns and the randomness in the data

Message + noise

Python Plotting

- Almost all types of plots done by `matplotlib` in Python
- You use "`import matplotlib.pyplot as plt`" to import all the functionalities
- Always follow the following steps
  1. Prepare data: either loading data from file, processing and computing
  2. Define a drawing window: size, subplots etc (or use the default setting by `plt.plot()`)
  3. Use the main plotting functions `plot` and/or `scatter` etc, depending on what plots are needed
- Please join the lab for training
- See a simple example in `Lecture02_Example00.py`
- Read some examples online

Time Series Plots

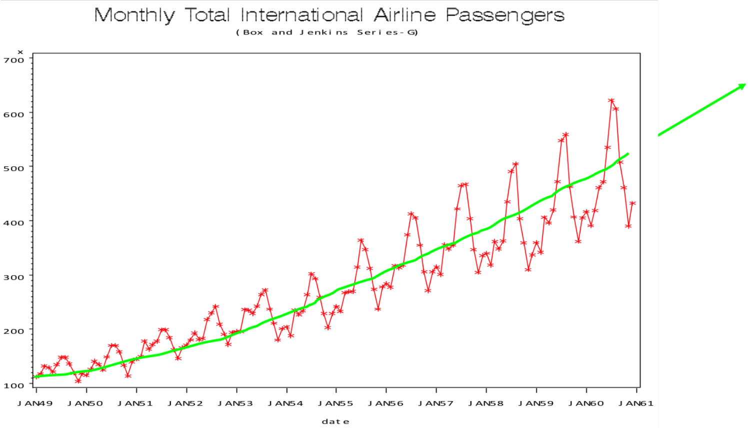


Time Series Components

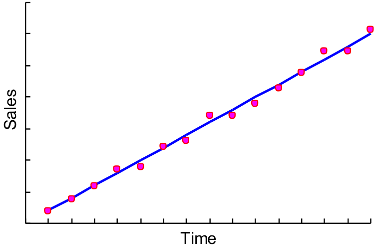
See an example in `Lecture02_Example01.py`

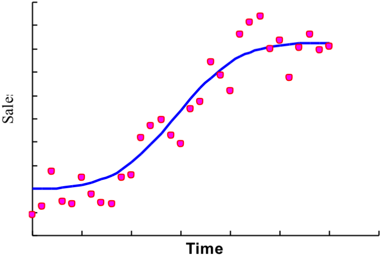
Trend

- Reflects the long-run growth or decline in the time series

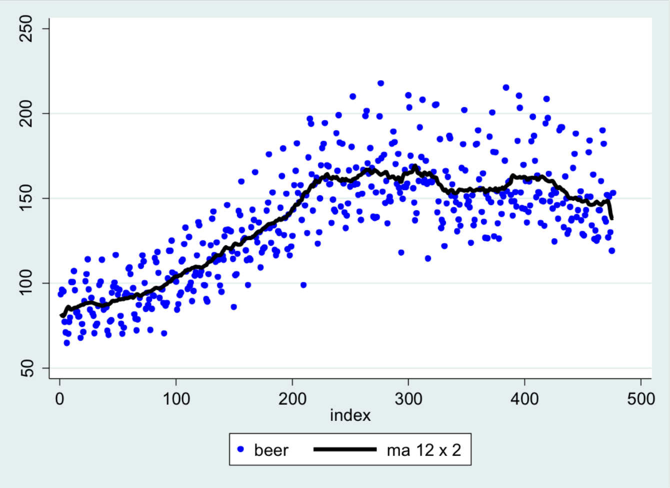


The Trend Component





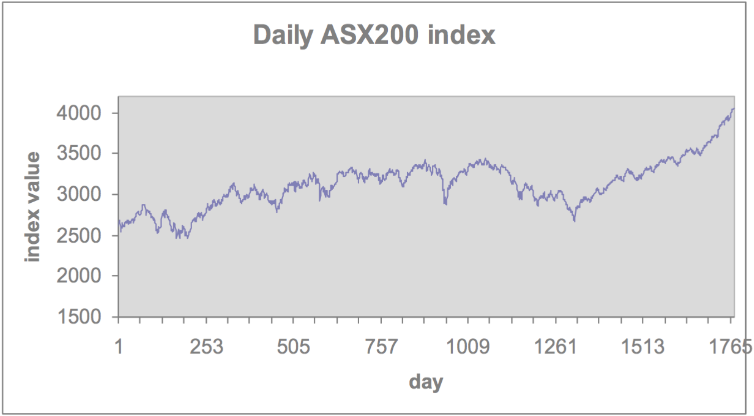
What type of trend do we have in this graph below? Is the trend useful for forecasting?



MUST be careful! May or MAY NOT continue

Cycle

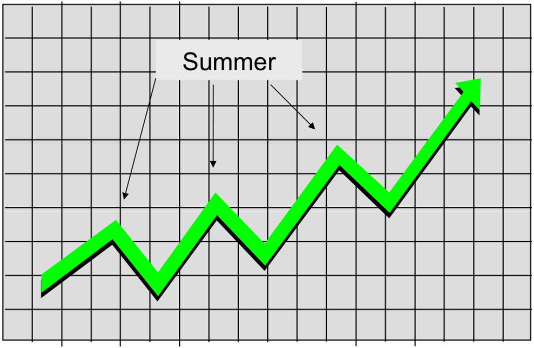
- Slow rises and falls that are not in a regular repeating pattern, no fixed period



Cycle

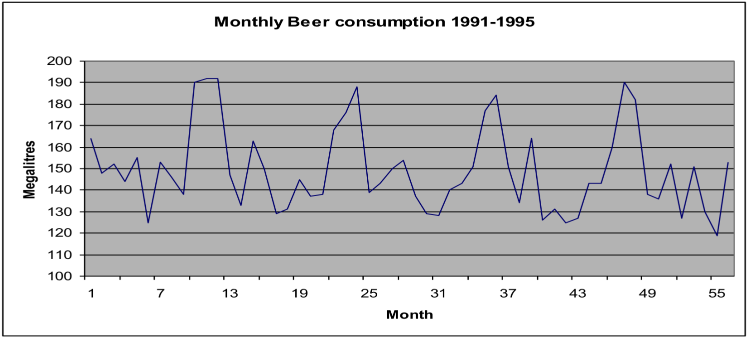
- Thus, usually very difficult to capture in a model!
- cycle assumed negligible or unforecastable
- trend-cycle analysed together
- cycle modelled by
    - ARIMA (later)
    - Regression
    - Constant

Seasonal

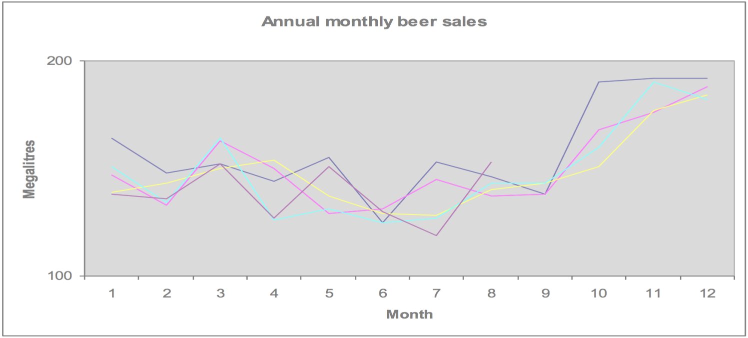


- Rises and falls that ARE in a regular repeating pattern
- There is seasonal period, denoted by M

What is the seasonal pattern here and why?

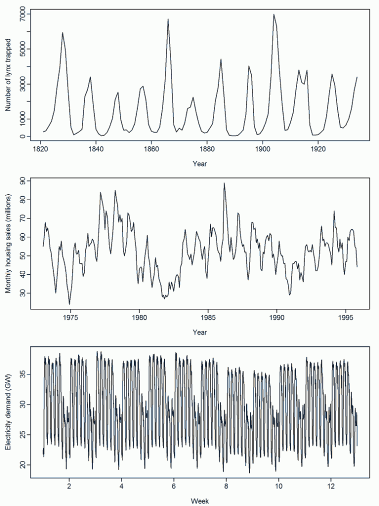


Time Series – Seasonal Plot



See an example in `Lecture02_Example02.py`

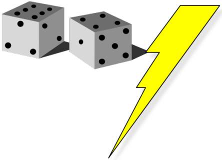
Difference between Seasonal and Cycle



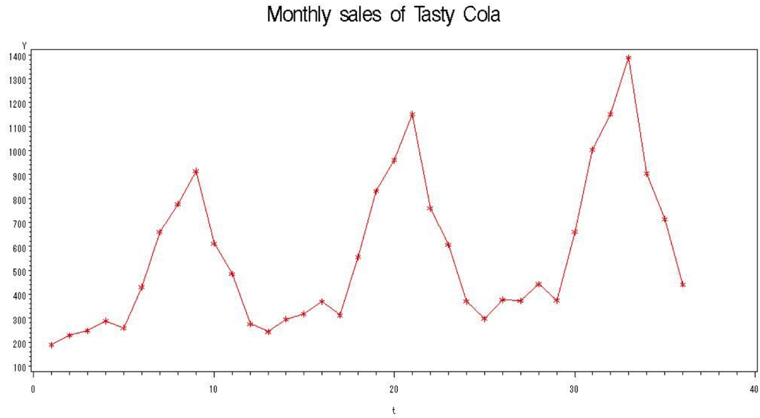
Source: `http://robjhyndman.com/hyndsight/cyclicts/`

Irregular fluctuations

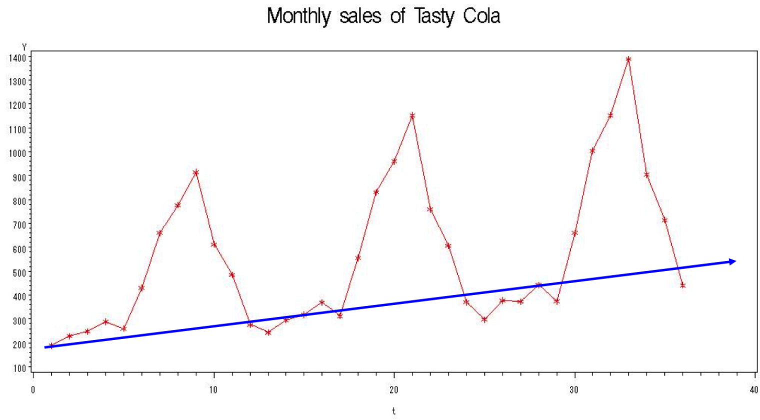
- follow no pattern
- assumed unexplainable
- Might be 'unusual' events: earthquakes, accidents, hurricanes, wars, strikes
- OR just random variation i.e. noise!



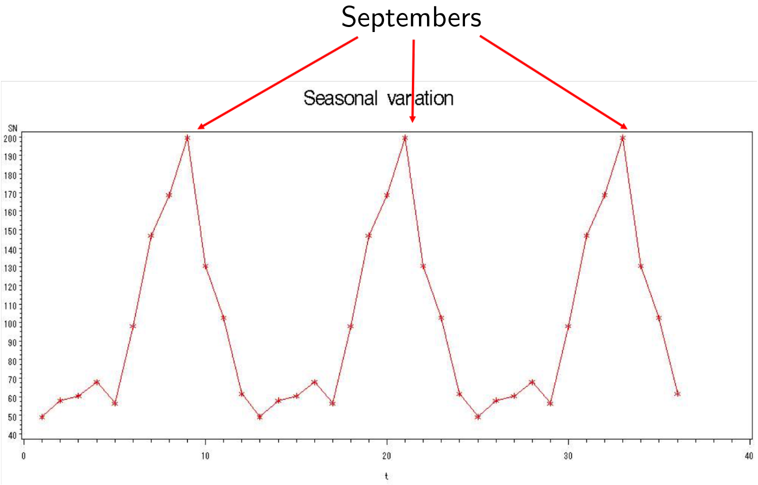
The Discount Soda Shop wants to forecast monthly Tasty Cola sales (in hundreds of cases)



The Upward Trend

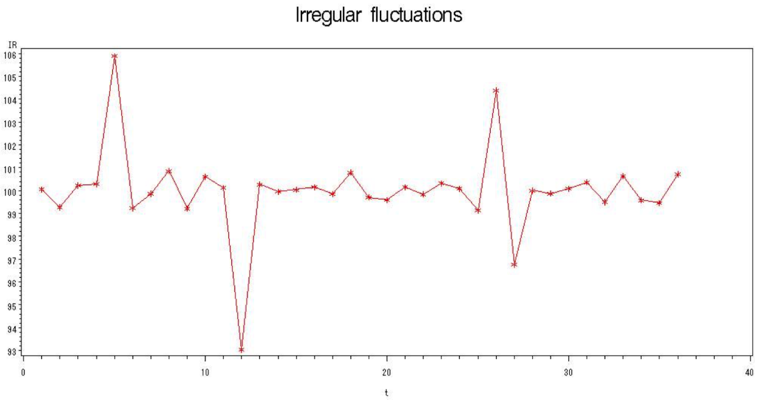


Seasonal Patterns:



$$M = 12$$

Irregular Fluctuations



Notation

- We denote a time series of length T (or N) as

$$\mathcal{Y} = \{Y_1, Y_2, ..., Y_T\}$$

- The observation and estimate at time point t are, respectively, denoted by

$$Y_t \quad \text{and} \quad \widehat{Y}_t$$

- The estimate error (residual)

$$e_t = Y_t - \widehat{Y}_t$$

- The seasonal period/frequency by M and the general seasonal index (for numbering) m (taking one value of 1, ..., M)

Notation for Seasonal Component

- We use $\mathcal{S} = \{S_1, S_2, ..., S_T\}$ for a seasonal series of period M where S_t is a generic notation. For example

$$\mathcal{S} = \{S_1, S_2, \cdots, S_{100}\} = \{-1, 2, 4, -6, -1, 2, 4, -6, \cdots, -1, 2, 4, -6\}$$

  where $M = 4$. Hence from the notation we know

$$S_1 = -1, S_2 = 2, S_3 = 4, S_4 = -6, S_5 = -1, S_6 = 2, \cdots, S_{99} = 4, S_{100} = -6$$

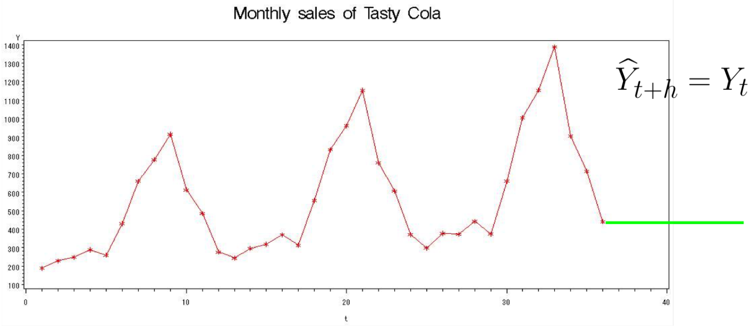
- Sometimes we use $\{\overline{S}_m\}$ $(m = 1, 2, \cdots, M)$ to denote all the M values. For example

$$\overline{S}_1 = -1, \overline{S}_2 = 2, \overline{S}_3 = 4, \overline{S}_4 = -6,$$

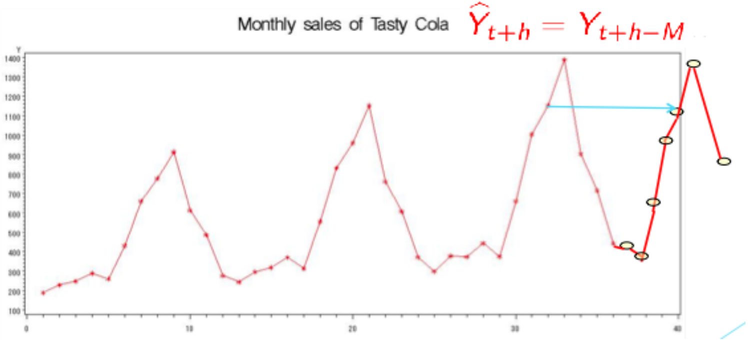
- Hence for any time t, S_t should be one of M (different) $\overline{S}_m$ $(m = 1, 2, \cdots, M)$
- Understanding this notation is critical for the Decomposition Methods next week, where $\overline{S}_m$ are called seasonal index.

Naïve Method: Most Recent Value

- This is the BASE model to which all forecast models should be compared



Seasonal Naïve Method: Most Recent Season's Value



Drift Method

- A variation on the naïve method, allowing the changes in the forecasts over time
- The amount of change over time, called `drift`, is the average change seen in the historical data.
- Give a (historical) time series

$$\mathcal{Y} = \{Y_1, Y_2, \dots, Y_T\}$$

- The drift method defines the forecast for the time point $T + h$ as

$$\widehat{Y}_{T+h} := Y_T + \frac{h}{T-1} \sum_{t=2}^{T} (Y_t - Y_{t-1})$$

  i.e., adding (h) times of the average change to the most recent observation Y_T.
- It can be proved that

$$\widehat{Y}_{T+h} = Y_T + h\left(\frac{Y_T - Y_1}{T-1}\right)$$

- This is equivalent to drawing a line between the first and last observation, and use that line to forecast for times after T.

Moving Average Methods

- MA is often used to smooth out noise and reveal the other 3 components:
- Smoothing (a 5 period centred MA)

$$\widehat{Y}_t = \frac{Y_{t+2} + Y_{t+1} + Y_t + Y_{t-1} + Y_{t-2}}{5}$$

- Example: Smoothing a time series

$$\mathcal{Y} = \{Y_1, Y_2, Y_3, Y_4, Y_5, Y_6, Y_7, Y_8, Y_9, Y_{10}\}$$

- The time point we can start smooth is $t = 3$, and the last time point we can smooth is $t = 8$. The smoothed time series is

$$\widehat{\mathcal{Y}} = \{-, -, \widehat{Y}_3, \widehat{Y}_4, \widehat{Y}_5, \widehat{Y}_6, \widehat{Y}_7, \widehat{Y}_8, -, -\}$$

  losing two values on both sides, where we calculate them as

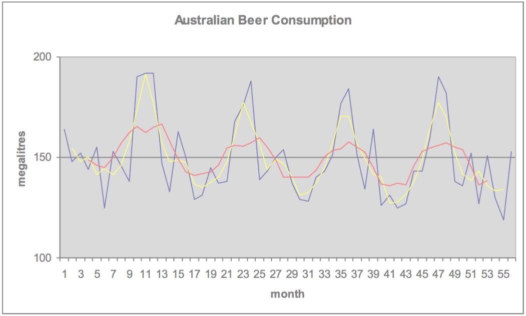
$$\widehat{Y}_3 = \frac{Y_1 + Y_2 + Y_3 + Y_4 + Y_5}{5}, \dots, \widehat{Y}_8 = \frac{Y_6 + Y_7 + Y_8 + Y_9 + Y_{10}}{5}$$

- Question: How to write the smoothing formula for a k (k is an odd integer) centred MA?

Selecting k

反应灵敏

- Heavier smoothing VS Responsiveness
- Useful to compare results with different k
- What will happen here if $k = 12$ or higher?
- Demo in `Lecture02_Example03.py`



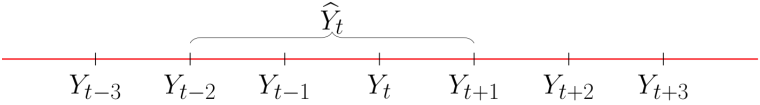
— MA 3 — MA 7

Even Order Moving Average: CMA-$(2k)$

- Odd Order MAs are symmetrically centred (i.e. Centred MA-5 or CMA-5 or simply MA-5)

$$\widehat{Y}_t = \frac{Y_{t+2} + Y_{t+1} + Y_t + Y_{t-1} + Y_{t-2}}{5}$$

- Consider possible MA-4. Which way?



$$\widehat{Y}_t = \frac{Y_{t-2} + Y_{t-1} + Y_t + Y_{t+1}}{4}$$

Even Order Moving Average: CMA-($2k$)

- Odd Order MAs are symmetrically centred (i.e. Centred MA-5 or CMA-5 or simply MA-5)

$$\widehat{Y}_t = \frac{Y_{t+2} + Y_{t+1} + Y_t + Y_{t-1} + Y_{t-2}}{5}$$

- Consider possible MA-4. Which way?

$$Y_{t-3} \quad Y_{t-2} \quad \underbrace{Y_{t-1} \quad Y_t \quad Y_{t+1} \quad Y_{t+2}}_{\widehat{Y}_t} \quad Y_{t+3}$$

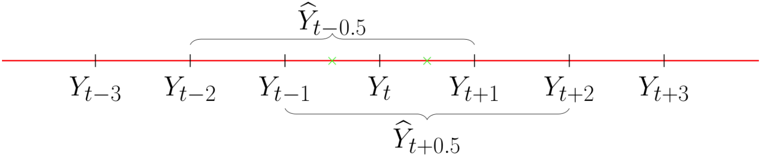
$$\widehat{Y}_t = \frac{Y_{t-1} + Y_t + Y_{t+1} + Y_{t+2}}{4}$$

Even Order Moving Average: CMA-$(2k)$

- Odd Order MAs are symmetrically centred (i.e. Centred MA-5 or CMA-5 or simply MA-5)

$$\widehat{Y}_t = \frac{Y_{t+2} + Y_{t+1} + Y_t + Y_{t-1} + Y_{t-2}}{5}$$

- Work out the half time smoothing, then average them



$$\widehat{Y}_{t-0.5} = \frac{Y_{t-2} + Y_{t-1} + Y_t + Y_{t+1}}{4}; \; \widehat{Y}_{t+0.5} = \frac{Y_{t-1} + Y_t + Y_{t+1} + Y_{t+2}}{4}$$

$$\widehat{\widehat{Y}}_t = \frac{\widehat{Y}_{t-0.5} + \widehat{Y}_{t+0.5}}{2} = \frac{1}{8} Y_{t+2} + \frac{1}{4}(Y_{t+1} + Y_t + Y_{t-1}) + \frac{1}{8} Y_{t-2}$$

Summary

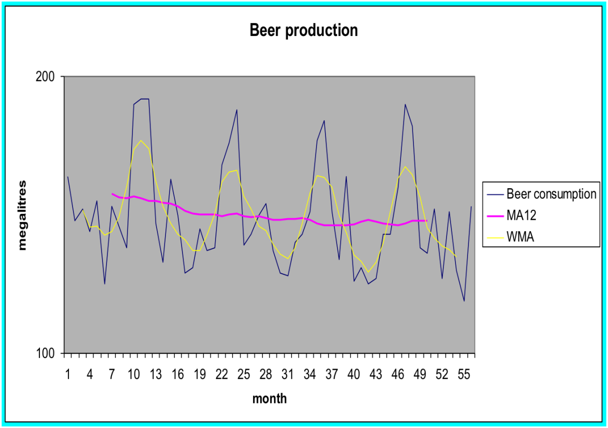
- Odd Order MAs are symmetrically centred.
- Even Order MAs are constructed in two-layer MA which are centred.
- We call them all the Centred MA, denoted by CMA-(M) where M is either odd number or even number (When M is odd number, simply denoted as MA-M)
- When M is even number, in the construction of CMA-M, we do two half centred averages at $t - 0.5$ and $t + 0.5$ first, then take the average of two half centred averages.
- Similarly we can conduct many layers even with different orders. For example:
  $m \times n$-MA means we will do m (C)MA-n's, then take the average over m (C)MAs.
- Derive the formula for 3×5-MA.

Example: Quarterly product sales

- Notice we lose 2 data points at the beginning of the series and 2 data at the end

t	Y_t	$\widehat{Y}_t$	$\widehat{\widehat{Y}}_t$	St
1	897			
2	476			
2.5		564.2		
3	376		573.1	0.656
3.5		582		
4	509		588.625	0.865
4.5		595.25		
5	967		599.125	1.614
5.5		603		
6	529		585.75	0.903
6.5		568.5		
7	407		558.125	0.729
7.5		547.75		
8	371		532.5	0.697
8.5		517.25		
9	884		504	1.75
9.5		490.75		
10	407			
11	301			

When k increase, we lose more data on both sides



See demo in `Lecture02_Example03.py`

Four Minutes Exercise

- Show that the formula for a CMA-6 is

$$\widehat{Y}_t = \frac{1}{12}(Y_{t+3} + Y_{t-3}) + \frac{1}{6}(Y_{t+2} + Y_{t+1} + Y_t + Y_{t-1} + Y_{t-2})$$

- The formula for a CMA-k (k is an even number) is

$$\widehat{Y}_t = \frac{1}{2k}(Y_{t+k/2} + Y_{t-k/2}) + \frac{1}{k}\left(Y_{t+k/2-1} + Y_{t+k/2-2} + \cdots \right.$$
$$\left. + Y_t + \cdots + Y_{t-k/2+2} + Y_{t-k/2+1}\right)$$

  Write the formula for $k = 8$

- Or we write it as the CMA-$(2k)$,

$$\widehat{Y}_t = \frac{1}{4k}(Y_{t+k} + Y_{t-k}) + \frac{1}{2k}\left(Y_{t+k-1} + Y_{t+k-2} + \cdots \right.$$
$$\left. + Y_t + \cdots + Y_{t-k+2} + Y_{t-k+1}\right)$$

  In this form, $k = 3$ corresponds to CMA-6.

- Prove the formula for the CMA-$(2k)$ (Homework!)

Moving Average for Forecasting

- Forecasting (using k pasts)

$$\widehat{Y}_{t+1} = \frac{Y_t + Y_{t-1} + Y_{t-2} + \cdots + Y_{t-k+1}}{k}$$

- For example, MA-5 for forecasting

$$\widehat{Y}_{t+1} = \frac{Y_t + Y_{t-1} + Y_{t-2} + Y_{t-3} + Y_{t-4}}{5}$$

- At $t = 5$, we forecast for the time $t + 1 = 6$ as

$$\widehat{Y}_6 = \frac{Y_5 + Y_4 + Y_3 + Y_2 + Y_1}{5}$$

which is the same as the smoothed value $\widehat{Y}_3$ at time $t = 3$ via MA-5 Smoothing.

Moving Average for Forecasting

- In terms of algorithm, we can do a MA-5 smoothing, and shift THREE units time, then we have a one-step ahead forecast of MA-5
- This can be shown by the following pattern, (example for MA-5),
  Smoothing:

$$
\begin{array}{cccccccc}
Y_1, & Y_2, & Y_3, & Y_4, & \cdots, & Y_{T-2}, & Y_{T-1}, & Y_T \\
 & & \widehat{Y}_3, & \widehat{Y}_4, & \cdots, & \widehat{Y}_{T-2} & &
\end{array}
$$

Forecasting, shifting THREE units time

$$
\begin{array}{cccccccccc}
Y_1, & Y_2, & Y_3, & Y_4, & Y_5, & Y_6, & \cdots, & Y_T & ??? \\
 & & & & & \widehat{Y}_3, & \cdots, & \widehat{Y}_{T-3}, & \widehat{Y}_{T-2}
\end{array}
$$

Note here $\widehat{Y}_3$ is the average of Y_1, Y_2, Y_3, Y_4, Y_5, ...

Weighted Moving Averages (WMA-5)

- Smoothing (A window-5 WMA)

$$\widehat{Y}_t = w_1 Y_{t+2} + w_2 Y_{t+1} + w_3 Y_t + w_4 Y_{t-1} + w_5 Y_{t-2}$$

where $\sum_{i=1}^{5} w_i := w_1 + w_2 + w_3 + w_4 + w_5 = 1.$

- Forecasting (WMA-k)

$$\widehat{Y}_{t+1} = w_1 Y_t + w_2 Y_{t-1} + w_3 Y_{t-2} + \cdots + w_{k-1} Y_{t-k+2} + w_k Y_{t-k+1}$$

where $\sum_{i=1}^{k} w_i = 1.$

\hat{Y_{6}}=w1*Y_{5}+w2*Y4+w3*Y3+w4*Y2+w5*Y1

Weighted Moving Averages (WMA): Examples

- Smoothing (A (symmetric) WMA-5)

$$\widehat{Y}_t = 0.15Y_{t+2} + 0.2Y_{t+1} + 0.3Y_t + 0.2Y_{t-1} + 0.15Y_{t-2}$$

- Smoothing (A CMA-4 is a WMA-5)

$$\widehat{Y}_t = \frac{1}{8}(Y_{t+2} + Y_{t-2}) + \frac{1}{4}(Y_{t+1} + Y_t + Y_{t-1})$$

More Examples

- Some popular sets of weights

WMA-5:

$\widehat{Y}_t = \frac{1}{9}(Y_{t+2} + Y_{t-2}) + \frac{2}{9}(Y_{t+1} + Y_{t-1}) + \frac{1}{3} Y_t$

WMA-7:

$\widehat{Y}_t = \frac{1}{15}(Y_{t+3} + Y_{t-3}) + \frac{2}{15}(Y_{t+2} + Y_{t-2}) + \frac{1}{5}(Y_{t+1} + Y_t + Y_{t-1})$

CMA-12 (WMA-13):

$$\begin{aligned}
\widehat{Y}_t =& \frac{1}{24}(Y_{t+6} + Y_{t-6}) + \frac{1}{12}(Y_{t+5} + Y_{t+4} + Y_{t+3} + Y_{t+2} \\
&+ Y_{t+1} + Y_t + Y_{t-1} + Y_{t-2} + Y_{t-3} + Y_{t-4} + Y_{t-5})
\end{aligned}$$

H9: Due to Henderson

$$\begin{aligned}
\widehat{Y}_t =& 0.33 Y_t + 0.267(Y_{t+1} + Y_{t-1}) + 0.119(Y_{t+2} + Y_{t-2}) \\
&- 0.01(Y_{t+3} + Y_{t-3}) - 0.0041(Y_{t+4} + Y_{t-4})
\end{aligned}$$

MA smoothing — some notes

- Smoothing can reveal structure and components more clearly than raw data.
- Later, smoothing will help us model and estimate these components.
- Weighted MA smoothing can be better than equal weights.
- If data have seasonal frequency M, a (C)MA-M smoother will remove the seasonal pattern to reveal trend and cyclic patterns. Recall our demo. Can you show me why?
- New approach is to learn weights for MA that most suits for a special kind of time series

The forecast error

- The forecast error for a particular forecast is:

$$e_t = Y_t - \widehat{Y}_t$$

- BUT SOMETIMES
  - directions are measured, e.g., Did Y go UP when forecast said it would go UP?
  - categories are forecast: In these cases errors are measured in terms of disagreement

$$d_t = \begin{cases} 0 & Y_t = \widehat{Y}_t, \\ 1 & Y_t \neq \widehat{Y}_t. \end{cases}$$

    Did actual 'land' in forecasted category (Election, sport result, positive profit for stock portfolio, portfolio BEAT market, etc.)
  - MUST have actual data to do this!

Measuring Forecast accuracy

- A forecast method is unbiased if:

$$E(e_t) = 0 \Longleftrightarrow E(Y_t) = E(\widehat{Y}_t)$$

which implies that:

$$\frac{1}{h}\sum_{t=T+1}^{T+h}(Y_t - \widehat{Y}_t) \approx 0.$$

- Is this a good criterion to assess forecast accuracy? WHY?

Measuring Size/Magnitude of Forecast Errors

- Mean Absolute Deviation (MAD)

$$\text{MAD} = \frac{1}{h} \sum_{t=T+1}^{T+h} |Y_t - \widehat{Y}_t|$$

MAD is average distance between actual and forecast, i.e. average forecast error.
and

- Mean Squared Error (MSE)

$$\text{MSE} = \frac{1}{h} \sum_{t=T+1}^{T+h} (Y_t - \widehat{Y}_t)^2$$

MSE is like a forecast variance, if forecasts are unbiased. RMSE is forecast standard deviation, just take square root of MSE.

- MAD and RMSE ($\sqrt{MSE}$) are in original units of data. MSE penalises a small number of large errors more than MAD.

Mean Absolute Percentage Error

$$\mathrm{APE}_t = \frac{|Y_t - \widehat{Y}_t|}{Y_t} \times 100$$

$$\mathrm{MAPE} = \frac{1}{h} \sum_{t=T+1}^{T+h} \mathrm{APE}_t$$

- Forecast errors are percentages of the actual data point, e.g. 10%. Very popular in business forecasting
- Knowing that MAPE is 10%, i.e. Average error is 10% can be more valuable than knowing it is e.g. 12 Megalitres (MAD or RMSE)
- Cannot be used if any $Y_t = 0$.

Which measure is best?

- MAD
  - same units as Y
  - Does not heavily penalise a small number of large errors
- MSE
  - Harder to interpret
  - Heavily penalises large errors
  - RMSE has same units as Y
- MAPE
  - measures percentage error
- ALL can be used simultaneously. Report measure(s) that decision maker/manager can BEST understand. Are one or two large errors highly UNDESIRABLE? Yes? Use MSE. No? Use MAD or MAPE

Which measure is best?

- For categorical or direction forecasting

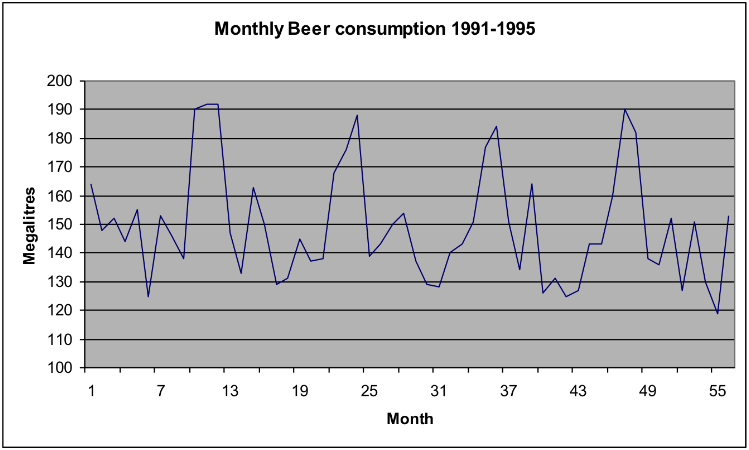
  Percentage agreement
  and/or
  Percentage disagreement

  are often used.
- e.g. how many elections has this model correctly forecast, compared to the total? Ans. $\frac{3}{4} = 75\%$
- Logit models are often used (ECMT2120). Can also compare forecasted probabilities with % occurrences e.g. Betting agencies might put odds of 2 to 1 on the Bulldogs beating Roosters. In games with these odds, did the favourite win (close to) 2 out 3 as forecasted?

Out-of-sample forecasting

- A realistic way of assessing a model's accuracy is to use a holdout set. i.e. Hold some data for testing forecasting
- E.G.: we hold out the last 8 months of a dataset.
- We use the previously observed data to produce forecasts for these 8 months

Example - adapted from MWH pg 44-45



Example - adapted from MWH pg 44-45

Time	Data	Forecast **Naive**	Error	Absolute error	Percentage error	Absolute percentage error
t	y_t					
50	138	182	-44.0	44.0	-31.9	31.9
51	136	138	-2.0	2.0	-1.5	1.5
52	152	136	16.0	16.0	10.5	10.5
53	127	152	-25.0	25.0	-19.7	19.7
54	151	127	24.0	24.0	15.9	15.9
55	130	151	-21.0	21.0	-16.2	16.2
57	119	130	-11.0	11.0	-9.2	9.2
56	153	119	34.0	34.0	22.2	22.2
		Mean:	-3.6	22.1	-3.7	15.9

Example - adapted from MWH pg 44-45

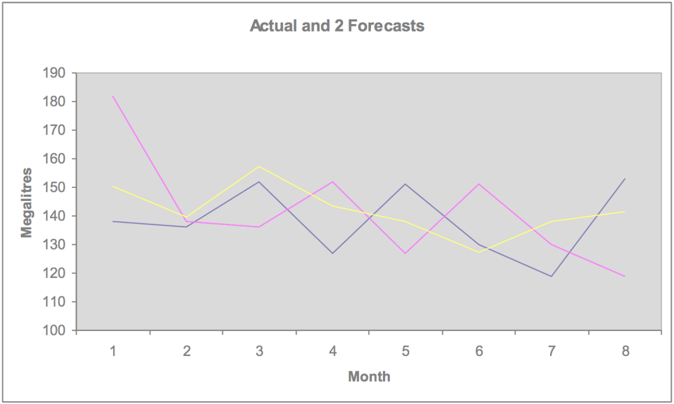
Time	Data	Forecast: **seasonal MA4**	Error	Absolute error	Percentage error	Absolute percentage error
t	y_t					
50	138	150.25	-12.25	12.25	-8.9	8.9
51	136	139.5	-3.5	3.5	-2.6	2.6
52	152	157.25	-5.25	5.25	-3.5	3.5
53	127	143.5	-16.5	16.5	-13	13
54	151	138	13	13	8.6	8.6
55	130	127.5	2.5	2.5	1.9	1.9
57	119	138.25	-19.25	19.25	-16.2	16.2
56	153	141.5	11.5	11.5	7.5	7.5

Mean: -3.72 10.47 -3.3 7.8

Comparison of forecast methods

Method	ME	MAD	MAPE	MSE	RMSE
Naive	**-3.6**	**22.1**	**15.9**	**639.4**	**25.3**
Seasonal	**-3.7**	**10.5**	**7.8**	**142.5**	**11.9**
Units	**Ml**	**Ml**	**%**	**Ml squared**	**Ml**

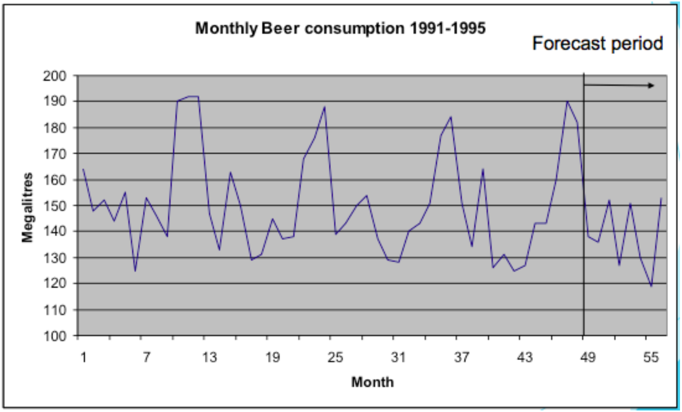
Comparison of forecasts



— Actual — Naïve — Seasonal

Multi-step Horizon Forecasting:

- Alternatively, we can investigate what happens if the last observations are not available at the moment of the forecast
- Consider monthly beer consumption in Australia



Monthly beer consumption in Australia

Time	Data	Forecast **Naive**	Error	Absolute error	Percentage error	Absolute percentage error
t	y_t					
50	138	**182**	-44	44	-31.88	31.88
51	136	**182**	-46	46	-33.82	33.82
52	152	**182**	-30	30	-19.74	19.74
53	127	**182**	-55	55	-43.31	43.31
54	151	**182**	-31	31	-20.53	20.53
55	130	**182**	-52	52	-40.00	40.00
57	119	**182**	-63	63	-52.94	52.94
56	153	**182**	-29	29	-18.95	18.95

Mean: -43.8 43.8 -32.7 32.7

Monthly beer consumption in Australia

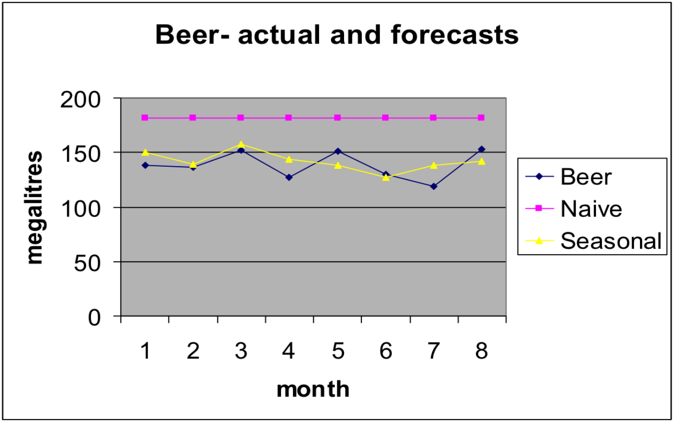
Time	Data	Forecast: **seasonal MA4**	Error	Absolute error	Percentage error	Absolute percentage error
t	y_t					
50	138	**150.25**	-12.25	12.25	-8.9	8.9
51	136	**139.5**	-3.5	3.5	-2.6	2.6
52	152	**157.25**	-5.25	5.25	-3.5	3.5
53	127	**143.5**	-16.5	16.5	-13	13
54	151	**138**	13	13	8.6	8.6
55	130	**127.5**	2.5	2.5	1.9	1.9
57	119	**138.25**	-19.25	19.25	-16.2	16.2
56	153	**141.5**	11.5	11.5	7.5	7.5

Mean: -3.72 10.47 -3.3 7.8

Comparison of forecast methods

Method	ME	MAD	MAPE	MSE	RMSE
Naive	-43.8	43.8	32.7	2056.5	43.3
Seasonal	**-3.7**	**10.5**	**7.8**	**142.5**	**11.9**
Units	**Ml**	**Ml**	**%**	**Ml squared**	**Ml**

Comparison of forecasts



Choosing a Forecasting Technique

A forecaster must consider :

- Time frame (time horizon): Statistical methods are generally MOST useful for the immediate to medium term.
  - Immediate $\Rightarrow \sim < 1$ month
  - Short-term $\Rightarrow \sim < 1 - 3$ months
  - Medium $\Rightarrow \sim < 3 - 18$ months
  - Long-term $\Rightarrow \sim > 18$ 18 months
- The cost of forecasting AND the accuracy desired**
- The availability of data
- The ease of operation and understanding
- What decisions will be made based on forecasts and by whom

Training and test sets

- It is important to evaluate forecasting accuracy using genuine forecasts. That is, it is invalid to look at how well a model fits the historical data. We can only determine the accuracy of forecasts by considering how well a model performs on new data
  - A model which fits the data well does not necessarily forecast well.
  - A near perfect t can always be obtained by using a model with enough parameters.
  - Over-fitting a model to data is as bad as failing to identify the systematic pattern in the data.
- Occam's Razor: The simpler the better

Training and test sets

- When choosing models, it is common to use a portion of the available data for fitting, and use the rest of the data for testing the model. Then the testing data can be used to measure how well the model is likely to forecast on new data.
- The size of the test set is typically about 20% of the total sample, although this value depends on how long the sample is and how far ahead you want to forecast. The size of the test set should ideally be at least as large as the maximum forecast horizon required.
- Time series references often call the training set the "in-sample data" and the test set the "out-of-sample data"

Cross-validation: For Cross-Sectional Data

- Select observation i (leave-one-out) for the test set, and use the remaining observations in the training set. Compute the error on the test observation.
- Repeat the above step for $i = 1, 2, ..., T$, where T is the total number of observations.
- Compute the forecast accuracy measures based on the errors obtained.

Cross-validation: Forward Chaining

- Suppose that we have a total of T observations and require At Least k observations to produce a reliable forecast. We implement the following steps
  1. Repeat the following steps for $i = 1, 2, ..., T - k$ where T is the total number of observations.
     - Select the observation at time $k + i$ for testing, and use the observations at times $1, 2, ..., k + i - 1$ to estimate the forecasting model.
     - Use the model to predict $\widehat{Y}_{k+i}$ for the time $k + i$ and compute the error on the forecast, $e_{k+i} = Y_{k+i} - \widehat{Y}_{k+i}$.
  2. Compute the forecast accuracy measures based on the errors obtained, $e_{k+1}, \ldots, e_T$.
- This procedure is sometimes referred to as expanding window forecasting and forward chaining

Example

- E.g., suppose $\mathcal{Y} = \{Y_1, ..., Y_{10}\}$ with $T = 10$ and we wish to assess a modelling method using At Least $k = 3$ previous observation. Then the previous forward chaining procedure will do the following, according the modelling method,
- Use Y_1, Y_2, Y_3 to build the model to predict Y_4, producing e_4
- Use Y_1, Y_2, Y_3, Y_4 to build the model to predict Y_5, producing e_5 (the algorithm may involves all the observations up to time 4)
- Use Y_1, Y_2, Y_3, Y_4, Y_5 to build the model to predict Y_6, producing e_6 (the algorithm may involves all the observations up to time 5)
- ... ...
- Use $Y_1, Y_2, Y_3, Y_4, Y_5, Y_6, Y_7, Y_8, Y_9$ to build the model to predict Y_{10}, producing e_{10} (the algorithm may involves all the observations up to time 9)
- Assess the overall errors $e_4, e_5, ..., e_{10}$ for the modelling method.

Cross-validation in Multi-step forecasts

- We follow an identical procedure to evaluate h step ahead forecasts.
  1. Repeat the following steps for $i = 1, 2, ..., T - k - h + 1$ where T is the total number of observations.
     - Select the observation at time $k + h + i - 1$ for testing, and use the observations at times $1, 2, ..., k + i - 1$ to estimate the forecasting model.
     - Use the model to predict $\widehat{Y}_{k+h+i-1}$ for the time $k + h + i - 1$ and compute the h-step error on the forecast for time $k + h + i - 1$, i.e., $e_{k+h+i-1} = Y_{k+h+i-1} - \widehat{Y}_{k+h+i-1}$
  2. Compute the forecast accuracy measures based on the errors obtained, $e_{k+h}, \ldots, e_T$
- When $h = 1$, this gives the same procedure as the previous one.
- Exercise: Let $T = 100$, $k = 4$ and $h = 3$, work out the above procedure.