QBUS6840 Lecture 02

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

Discipline of Business Analytics

The University of Sydney Business School

Lecture 1 recap

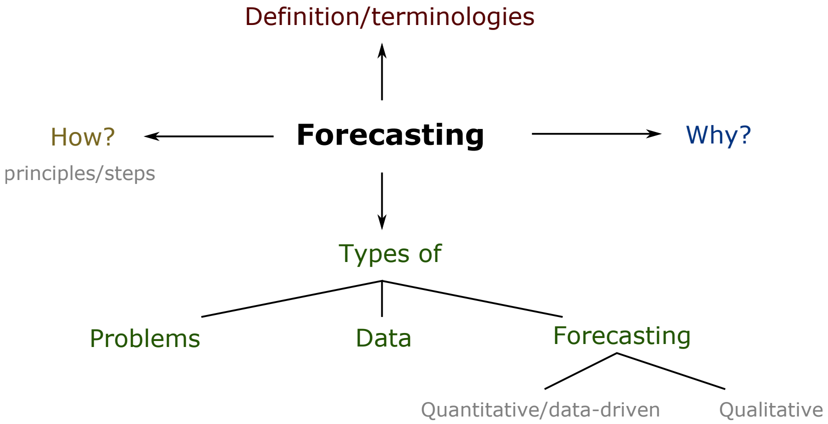


Table of contents

Readings

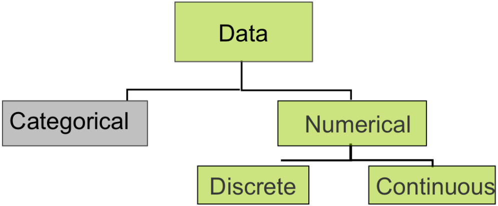
Online textbook Chapter 2: `otexts.com/fpp2/graphics.html`
and Chapter 3: `otexts.com/fpp2/simple-methods.html`

Outline

Data

- Data carry information/knowledge: categorical, 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 are collected in a passive way thanks to digital technology.

Data Types and Categories



- Structured Data
- Unstructured Data
- Text data
- Image Data
- Audio Data
- ... ...

Outline

Data Graphing

- Plotting data is the first step in any data analysis task, including predictive analytics task.
- It helps
  - Visualize many features of the data: trend, correlation relationship, etc.
  - Give insights into the data
- The type of the data determines the type of graphing technique
  - Two popular ones: time series plot and scatter plot

Data Graphing: 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://developers.google.com/chart/interactive/docs/gallery

Good graphs convey *both* patterns and the randomness in the data

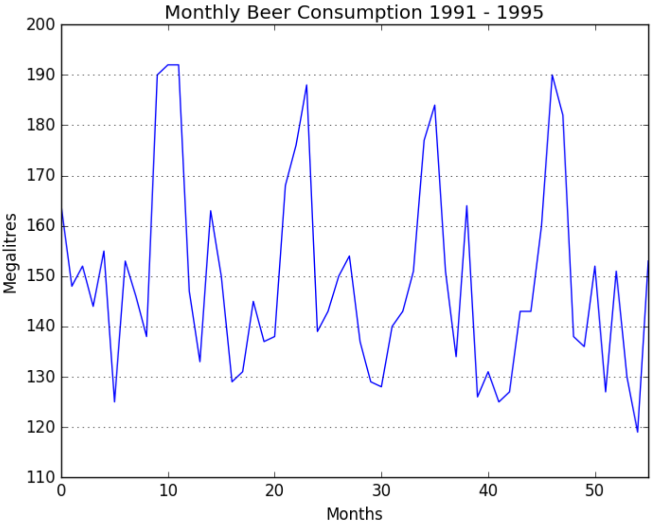
Message + noise

It's the data analyst's job to remove the noise using formal data analysis techniques.

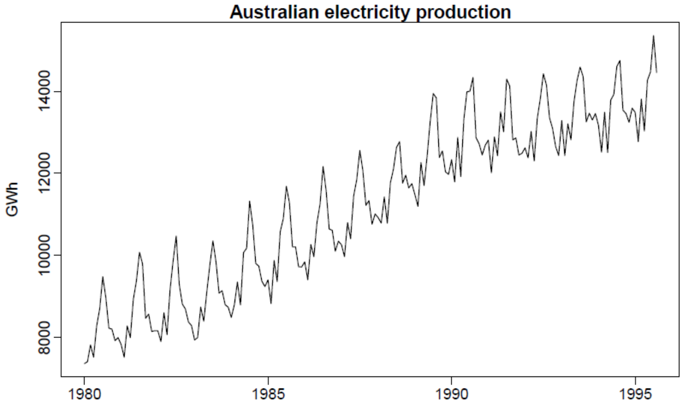
Python Plotting

- 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

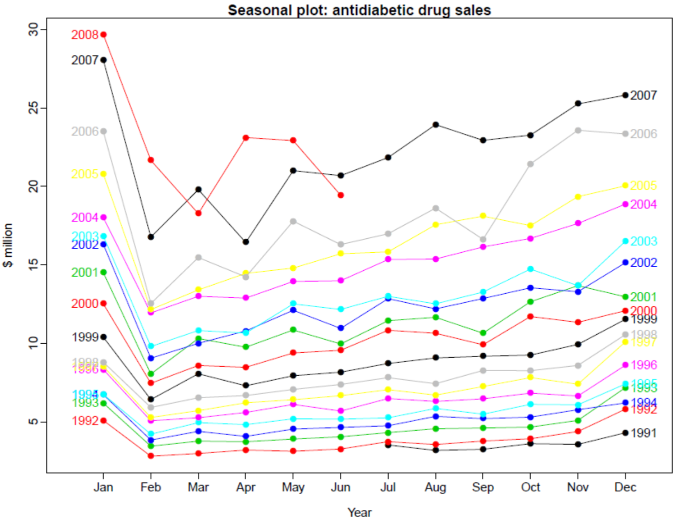
Time Series Plots



Time Series Plots



Time Series Plots



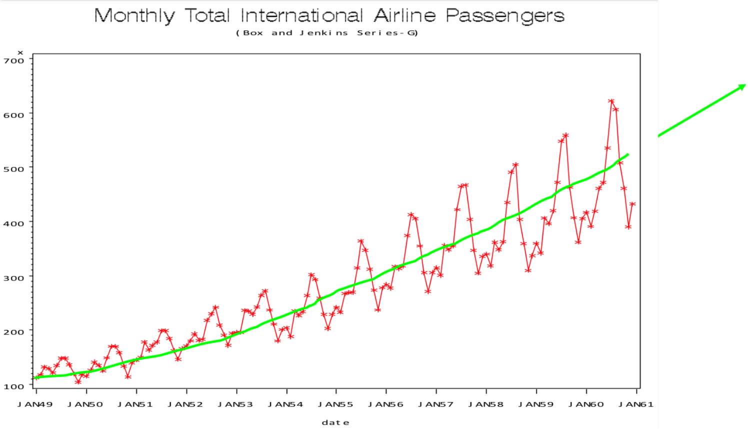
Outline

Time Series Components

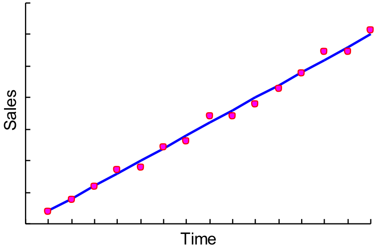
See an example in `Lecture02_Example01.py`

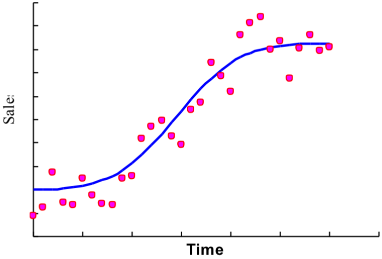
Trend

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

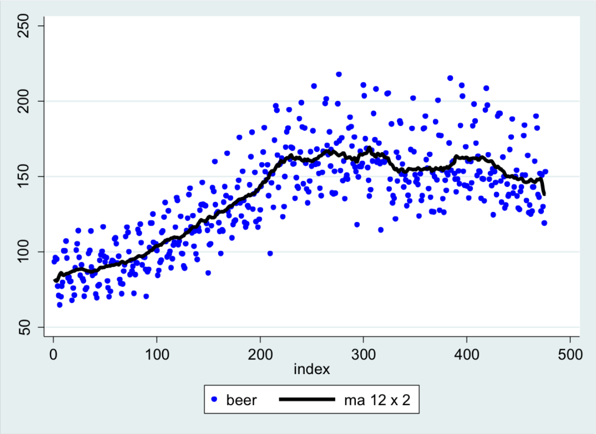


The Trend Component





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



MUST be careful! Trend may depend on the length of the observed time series

Cycle

- Slow rises and falls that are not in a regular repeating pattern, no fixed period
- often related to the "business cycle"

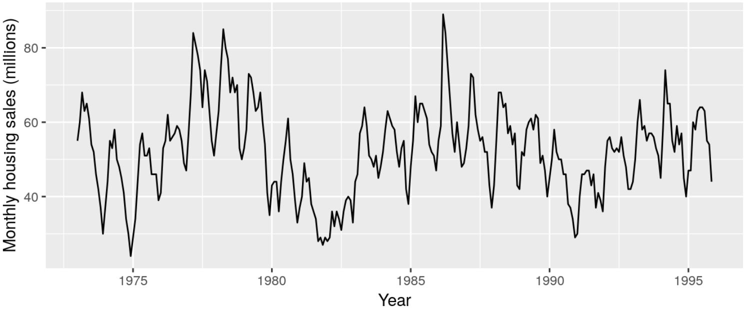
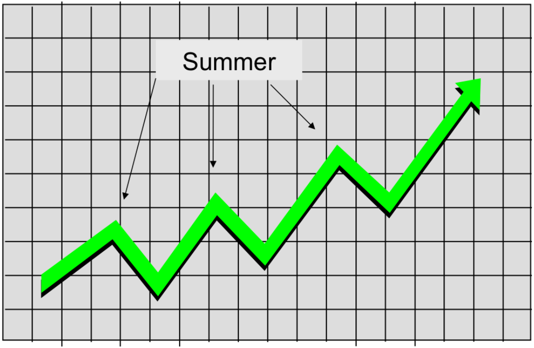


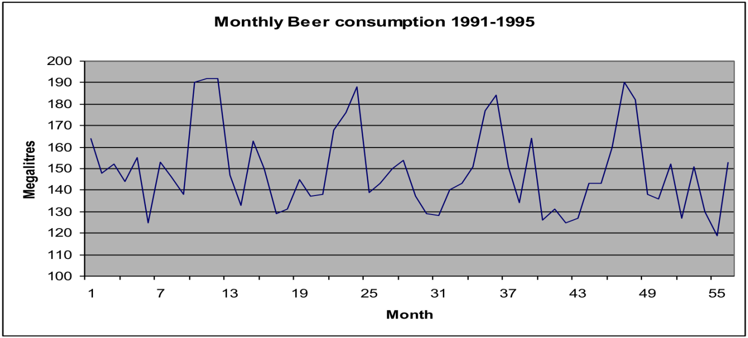
Figure: The data exhibit a cyclic pattern every 8-10 years

Seasonal

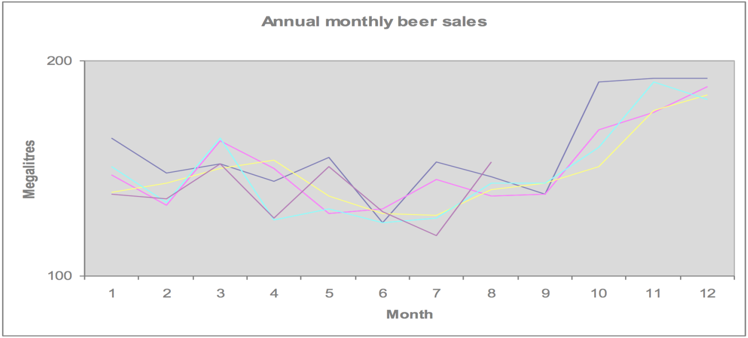


- Rises and falls that are in a regular repeating pattern, on a seasonal basis such as months of the year or days of the week.
- There is a fixed seasonal period/frequency, denoted by M

What is the seasonal pattern here and why?



Time Series – Seasonal Plot



See an example in `Lecture02_Example02.py`

Difference between Seasonal and Cycle

- Cycle: the rises and falls are not of a fixed frequency
- Seasonal: the rises and falls are associated with the calendar, and the frequency M is unchanging (every 12 months, 7 days, etc)

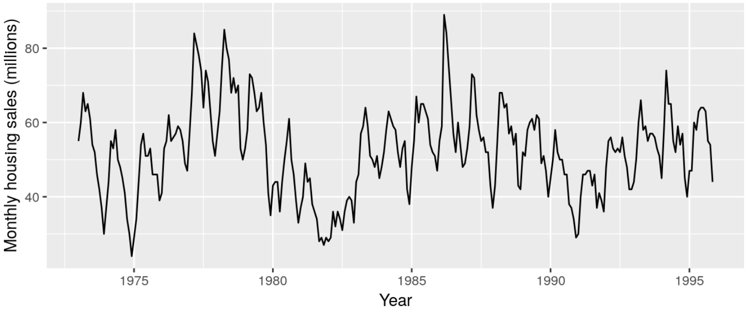
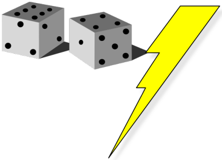


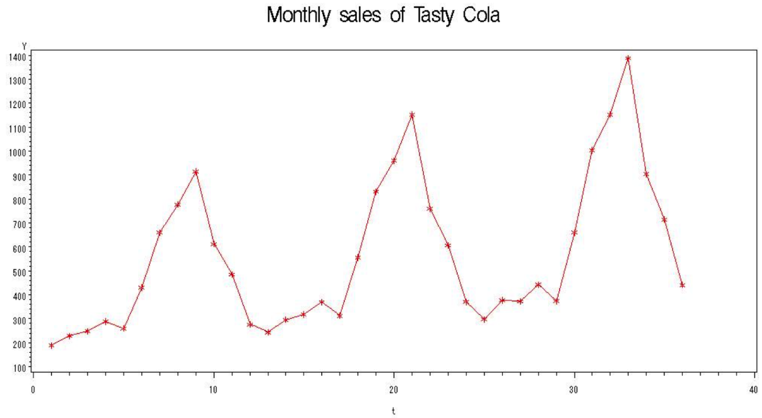
Figure: The data exhibit both a cyclic pattern and a seasonal pattern.

Irregular fluctuations

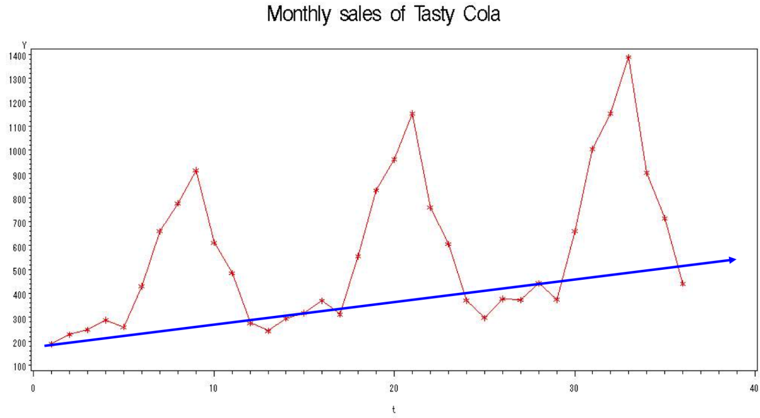
- follow no visualizable pattern, need statistical models to capture
- assumed unexplainable
- Might be 'unusual' events: earthquakes, accidents, hurricanes, wars, strikes
- OR just random variation i.e. noise!



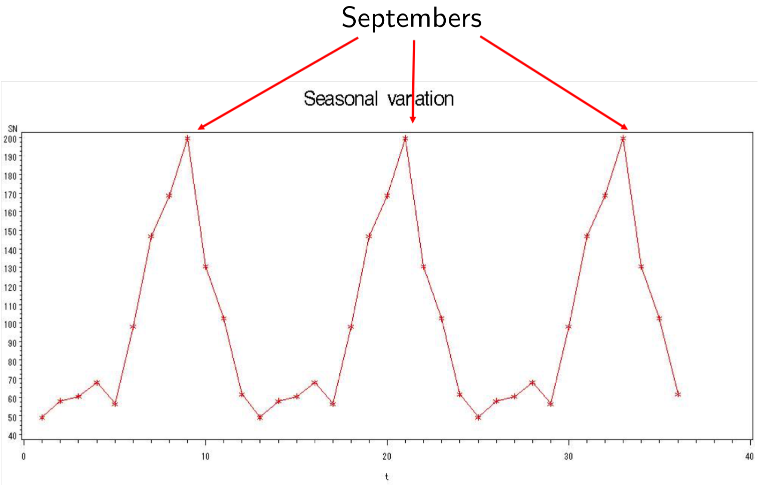
The Discount Soda Shop wants to forecast monthly Tasty Cola sales (in hundreds of cases)



The Upward Trend



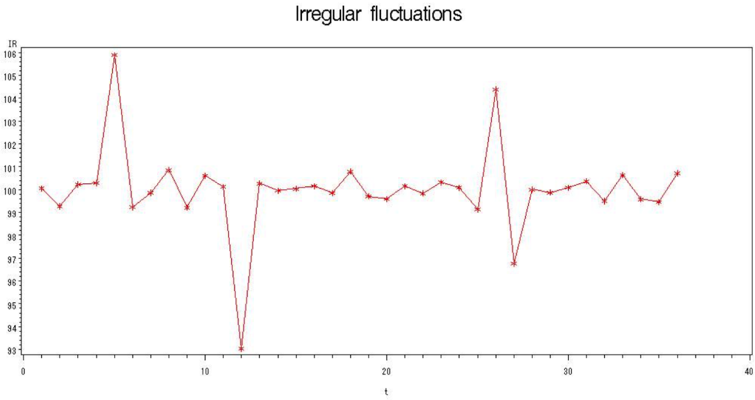
Seasonal Patterns:



$$M = 12$$

Irregular Fluctuations

... after the seasonal component is removed



Outline

Notation

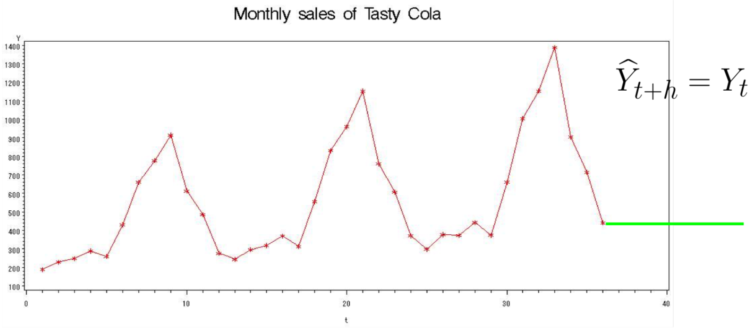
- We denote a time series of length T (or N) as

$$\mathcal{T} = \{Y_1, Y_2, ..., Y_T\}, \text{ or } Y_{1:T}$$

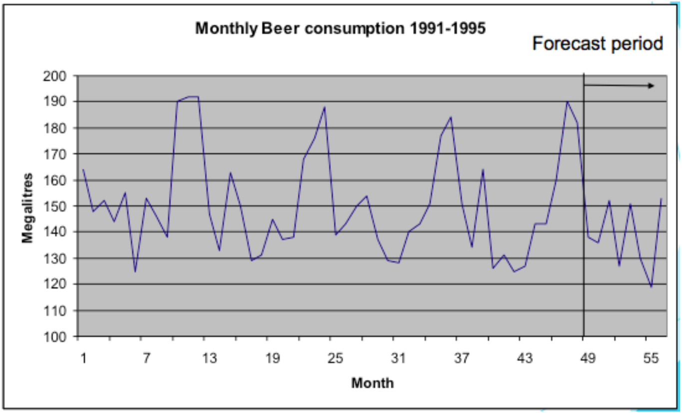
  where Y_t is the observation at time point t.
- The forecast of Y_{T+h} based on data $Y_{1:T}$ is denoted as $\widehat{Y}_{T+h|T}$, $h = 1, 2, ...$ is called horizon.
  - Called h-step-ahead forecast.
  - Sometimes, we simply write $\widehat{Y}_{T+h}$ for $\widehat{Y}_{T+h|T}$.

Naïve forecasting method: Most Recent Value

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



Monthly beer consumption in Australia



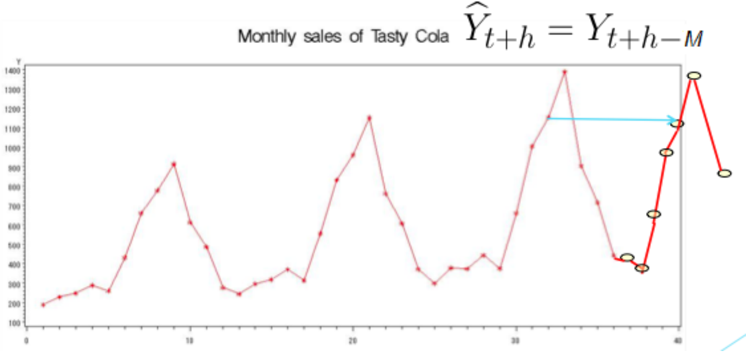
Monthly beer consumption in Australia: naive forecast

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

Seasonal Naïve Method: Most Recent Season's Value

$$\widehat{Y}_{t+h} = Y_{t+h-M}, \quad h = 1, 2, \ldots$$



Monthly beer consumption in Australia: seasonal naive forecast

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

Drift Method

- A variant of 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

$$\mathcal{T} = \{Y_1, Y_2, \ldots, Y_T\}$$

- 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.

Google stock price forecast

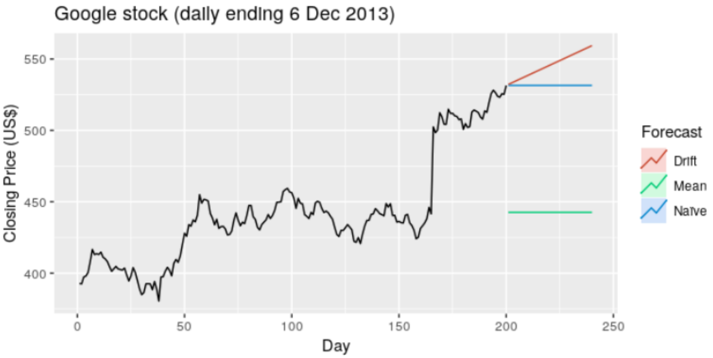


Figure 3.2: Forecasts based on 200 days of the Google daily closing stock price.

Stock prices are in general unpredictable, but their volatility is.

Outline

The forecast error

- Let $\widehat{Y_t} = \widehat{Y}_{t|t-1}$ be an one-step-ahead forecast of Y_t, based on data $Y_{1:t-1}$. The forecast error is the difference between Y_t and its forecast $\widehat{Y_t}$.
  - depends on how we define the "difference"
- For numerical data, the forecast error is often defined as

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

- For categorical data, the forecast error is measured in terms of disagreement

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

- MUST have actual data to compute errors!

Forecast accuracy measures

- 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? Answers: Should we choose method with sum of errors (i.e. average error) closest to 0? Desirable BUT what about SIGN of errors?

Accuracy measures

- 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.
- 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.
- Root Mean Squared Error (RMSE), the square root of MSE, is forecast standard deviation.
- MAD and RMSE are in original units of data. MSE penalises large errors more than MAD.

Accuracy measures

Mean Absolute Percentage Error

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

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

- Forecast errors are percentages of the actual data point, e.g. 10%. Very popular in business forecasting
- E.g., MAPE $= 10\%$ means that on average the forecast error is 10% of the data value.
  - sometimes, knowing that MAPE is 10%, say, can be more valuable than knowing it is e.g. 12 Megalitres (MAD or RMSE)
- Cannot be used if any $Y_t = 0$.

Monthly beer consumption in Australia: naive forecast

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: seasonal naive forecast

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

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?

- All the MAD, MSE, RMSE and MAPE are suitable for numerical data
- 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 (QBUS5001). Can also compare forecast probabilities with % occurrences e.g.
- There are other mearues of forecast accuracy - will be introduced along as we go.

Outline

Training and test sets

- An objective way of assessing a model's forecast accuracy is to use a test set for testing forecasting.
- It is often not enough 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 fit 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.

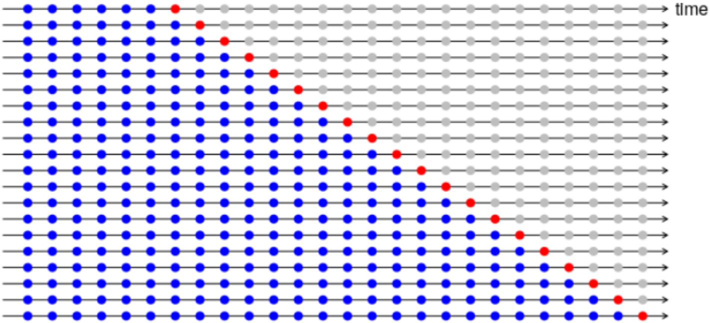
Cross-validation for time series (one-step-ahead forecast)

Given time series $Y_{1:T}$, use at least k observations for constructing forecasts

1. Repeat the following steps for $i = 1, 2, ..., T - k$
   - Select the observation Y_{k+i} for testing, and use the data $Y_{1:k+i-1}$ to estimate the forecasting model.
   - Use the model to compute the forecast $\widehat{Y}_{k+i}$ of Y_{k+i}.
   - Compute the forecast error e_{k+i}
2. Compute the forecast accuracy measures based on the errors obtained, $e_{k+1}, \ldots, e_T$.

Cross-validation for time series (one-step-ahead forecast)

The blue dots are training data, red dots are test data.



Cross-validation for time series (h-step-ahead forecast)

Given time series $Y_{1:T}$, use at least k observations for constructing forecasts.

1. Repeat the following steps for $i = 1, 2, ..., T - k - h + 1$
   - Select the observation $Y_{k+i+h-1}$ for testing, and use the data $Y_{1:k+i-1}$ to estimate the forecasting model.
   - Use the model to compute the h-step-ahead forecast $\widehat{Y}_{k+i+h-1}$ of $Y_{k+i+h-1}$.
   - Compute the forecast error $e_{k+i+h-1}$
2. Compute the forecast accuracy measures based on the errors obtained, $e_{k+h}, \dots, e_T$.

Cross-validation for time series (h-step-ahead forecast)

The Blue dots are training data, red dots are test data. $h = 4$.