QBUS6840 Lecture 01

Introduction

Professor Junbin Gao

The University of Sydney Business School

Outlines

- Unit House Keeping
- Introduction to Forecasting
- Prediction vs Forecasting
- Time Series vs Other Data
- Qualitative vs Quantitative
- Basic concepts of Forecasting

Objectives

- Understand the characteristics of time-series data in order to analyse real business data of this form
- Select and use an appropriate technique to predict the future behaviour of business variables of interest
- Be fluent in using computational tools to assist carrying out your analysis and generating visualisation
- Learn something about Python

Textbook

- Free Online Textbook
  - Rob J Hyndman and George Athanasopoulos, "Forecasting: principles and practice", an open access book available at www.otexts.org/fpp2/
- Main References
  - BOK: Bowerman, O'Connell and Koehler, "Forecasting, Time Series, and Regression," Duxbury Applied Series, 4th edition, 2005.
  - Comprehensive slides will be given weekly
  - Python online resources

Other References

- *Forecasting: Methods and Applications*, 3rd Edition, 1998 by Makridakis, Wheelwright, and Hyndman, Wiley, ISBN 0-471-53233-9. (MWH)
- *Business Forecasting*, 9th edition, 2008 by Hanke and Wichern, ISBN 0-13-500933-2.
- *Forecasting practice and Process for Demand Management*, 2006 by Levenbach H. and Cleary, J. P., ISBN0-534-26268-6, Thomson.
- *Applied Econometric Time Series*, by Walter Enders, 2nd Edition, Wiley 2004.
- *Introduction to Time Series and Forecasting*, 2nd edition, 2002 by Brockwell, P. J., and Davis, R. A. Springer-Verlag: New York, ISBN 0-387-95351-5.

Online Material

- Information on the Canvas https://canvas.sydney.edu.au
- Teaching Slides
- Tutorial material
- Canvas announcement (check the Canvas at least weekly)
- Some materials I found which may be of interest to you

The Team and Lecture Timetable

- Lectures (by Junbin Gao)
  - Fridays 11:00 - 13:00, ABS Auditorium (B2010)
  - Week 2 Friday 15:00 - 17:00, ABS Auditorium (B2010)
  - Week 8 Friday 18:00 - 20:00, ABS Auditorium (B2010)
- Lectures (by Chao Wang)
  - Tuesdays 18:00-20:00, Merewether Lecture Theatre 1 (Rm 131)
  - Week 8: Monday 18:00-21:00, Eastern Avenue Auditorium
- Consultation
  - Junbin Gao: Fridays 14:00 - 15:00, ABS 4085
  - Chao Wang: Tuesdays 14:30 - 15:30, ABS 4044.
- Tutorials:
  Run by PhD Students: Bingxin Zhou, Bingyuan Bai, Jiahao Xu, Suman Saha, Nick Nguyen, Jichan Kan and Boyang Zhang
  Attend one Stream only

Tutorial Timetable

There are in total 21 tutorial sessions

- Mondays 09:00-10:00: ABS Interactive Learning Studio 1090
- Mondays 09:00-10:00: Codrington Computer Laboratory 1 (H69.151)
- Mondays 12:00-13:00:JD Stewart Loxton Learning Studio 214
- Mondays 13:00-14:00: Education Learning Studio 229
- Mondays 17:00-18:00: New Law School Learning Studio 030
- Mondays 18:00-19:00: ABS Interactive Learning Studio 1090
- Mondays 20:00-21:00: ABS Interactive Learning Studio 1090
- Tuesdays 11:00-12:00: ABS Interactive Learning Studio 1090
- Tuesdays 11:00-12:00: New Law School Learning Studio 030
- Tuesdays 17:00-18:00: Codrington Computer Laboratory 1 (H69.151)
- Tuesdays 17:00-18:00: Codrington Computer Laboratory 5 (H69.133)
- Wednesdays 20:00-21:00: Codrington Computer Laboratory 1 (H69.151)
- Thursdays 16:00-17:00: Codrington Computer Laboratory 2 (H69.145)
- Thursdays 17:00-18:00: Codrington Computer Laboratory 2 (H69.145)
- Thursdays 18:00-19:00: Codrington Computer Laboratory 4 (H69.139)
- Fridays 08:00-09:00: Codrington Computer Laboratory 2 (H69.145)
- Fridays 10:00-11:00: Codrington Computer Laboratory 5 (H69.133)
- Fridays 11:00-12:00: Codrington Computer Laboratory 5 (H69.133)
- Fridays 12:00-13:00: Codrington Computer Laboratory 2 (H69.145)
- Fridays 13:00-14:00: Peter Nicol Russell Learning Studio 310
- Fridays 13:00-14:00: Codrington Computer Laboratory 2 (H69.145)

Tutor Consultation Timetable

- Nghia Nguyen (nghia.nguyen@sydney.edu.au):
  Mondays, 11am-12pm H69 Room 215
- Jihao Kan (jkan9151@uni.sydney.edu.au):
  Tuesdays, 13:00-14:001 H69 Room 215
- Jiahao Xu (jixu7952@uni.sydney.edu.au):
  Tuesdays, 16:00-17:00 H69 Room 215
- Boyan Zhang (bzha8220@uni.sydney.edu.au):
  Tuesdays, 18:00-19:00 ABS 2250
- Bingxin Zhou (bzho3923@uni.sydney.edu.au):
  Thursdays, 14:00-15:002 H69 Room 215
- Suman Saha (ssah4473@uni.sydney.edu.au):
  Thursdays, 16:00-17:00 H69 Room 215
- Mingyuan Bai (mbai8854@uni.sydney.edu.au):
  Fridays, 14:00-15:002 H69 Room 215

Assessments

- Homework (15%): due 12 April 2019
  Submission Cover Sheet on Canvas: Only your SID please (for anonymous marking)
- Group Assignment (30%): due 24 May 2019
  Submission Cover Sheet on Canvas: Only your group Number and Member SIDs please (for anonymous marking)
- Mid-semester Exam (20%): 2 Hours
  Week 8: Time TBA.
- Final Exam (35%): 3 Hours
  Final Exam Period
- Academic Honesty (compulsory by Week 4)

Academic Honesty Module

- Academic Honesty module may commence from Week 2
- All undergraduate and postgraduate students commencing their degrees
- Required to complete this new online module
- Students who must undertake the module should go `http://canvas.sydney.edu.au`
- Students must achieve 100% in all tests before the new module considers the student to have passed the module
- It takes around one hour to complete the module.
- Students must pass the Academic Honesty, or an Absent Fail (AF) grade will be given for the entire unit

Plagiarism

- Be careful to avoid Plagiarism
- What, Why and How: https://sydney.edu.au/students/academic-dishonesty-and-plagiarism.html
- Plagiarism means presenting work that is not your own without acknowledging the original source of the work. It doesnt matter whether you do this on purpose or accidentally.
- Never copy other stuff
- Turnitin system compares your assignment reports with other assignments, reports, journals etc (almost everything on the internet)

Group Project

- Each group consisting of 5 students (or 4 or 6 for a few groups)
- Groups will be formed in a self-choice way on Canvas. Please register a group with your favourite fellow students on Canvas, not necessarily in the same classes or tutorial streams
- Any remaining students will be randomly grouped, or added to some formed groups
- A group leader of each group will be automatically assigned by Canvas
- Attend group activities and complete the Group Project as a team
- Please refer to group project materials

Communication

- Emails the best option
  {junbin.gao,chao.wang}@sydney.edu.au
- Send me or the tutors email from your USyd official email address, don't send me emails from your work email addresses or other personal email addresses
- Always mention your full name and student ID

The Software for the Unit

- Python
  - Free and Works on PCs, Mac, Unix/Linux
  - Contains statistical modelling, visualisation and programming environment
  - Can be used for almost all models to be discussed in this class
  - `pandas`, `statsmodel`, `keras` packages needed for this course
- Matlab (not supported by us)
  - Licensed: USyd provides a license to enrolled students and can be downloaded and installed on personal computers, download it from https://sydney.edu.au/students/forms/student-it/matlab-use-agreement.html
  - Contains statistical modelling, visualisation and programming environment
  - Powerful Statistics Toolbox and Time Series Toolbox
- However excel is not enough to complete most data mining and machine learning tasks
- Most of these software are available in labs.

Some Tips

- You become a better problem solver by solving problems. Focus your efforts on the assignments and tutorials.
- High-quality feedback is available throughout the course so that you get the most out of QBUS6840. Ask for it.
- Ask questions in the lectures, consultation times, and by e-mail.
- Discuss the materials with your classmates/colleagues. Ask for helps.
- Look for answers on Google too, especially about programming issues.
- We focus on technical materials, but remember the profile of the data scientist: communication and data visualisation skills are very important.

An Introduction to Forecasting

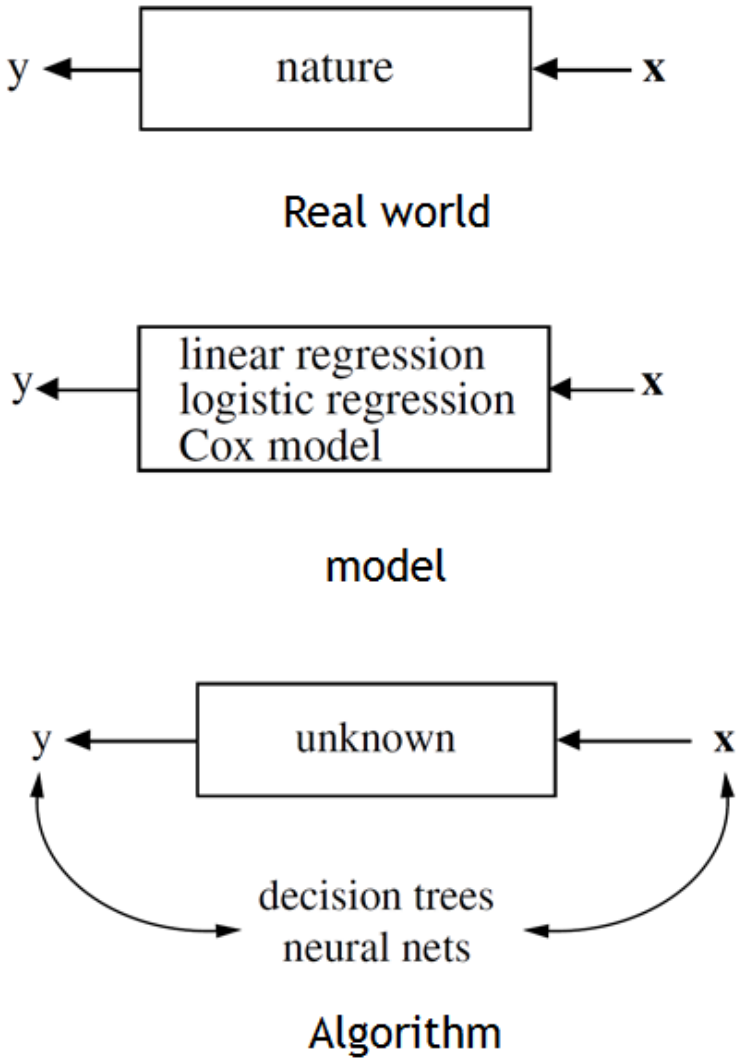
- Reading: https://otexts.com/fpp2/intro.html and https://otexts.com/fpp2/judgmental-adjustments.html
- Some useful terminology
- What exactly are we doing in the course? And what are we not doing?
- Difference: QBUS6810 (Statistical Learning and Data Mining), QBUS6830 (Financial Time Series and Forecasting), QBUS6850 (Machine Learning for Business)
- What is forecasting? Why forecasting?
- Overview of forecasting methods
- Principles of forecasting
- Key themes

Defining predictive modeling

- Predictive modeling: the process of developing a mathematical tool or model that generates an accurate prediction.
- Sometimes we only care about the predictions themselves. But I would add that it is often essential to establish and communicate the uncertainty in the predictions, as this is often important to assist decision making.
- Prediction vs Interpretation.

Models and algorithms

- Statistical modeling came out of statistics departments, machine learning algorithms came out of computer science departments. Certain methods and techniques are part of both. We can and should use both approaches.
- Breiman (2001) proposed the visualisation on the right: Two cultures



Types of problems and data

- Regression problems and classification problems.
- Time series (major part of the course) and cross-sectional (a few introduction) data.

Regression problems

- In regression problems we want to predict a numerical outcome.
  - How many copies will this book sell?
  - What will inflation be next month?
  - How much will my house sell for in the current market?
  - How much is this customer going to spend in my website today, given that he is going to purchase something?
  - How many tourists are going to visit NSW within the next years?
- Examples of regression models/algorithms: linear regression, penalized linear regression, partial least squares, neural networks, regression trees, etc.

Classification Problems

- Classification involves mapping your data points into a finite set of labels or the probabilities for each label. Some examples:
  - "Will someone click on this ad?" 0 or 1 (no or yes)
  - "What is this news article about?" politics, sports, culture ...
  - "What number is this? (image recognition)" 0, 1, 2, ...
  - "Is this message spam" 0 or 1
  - "Is this transaction fraudulent?" 0 or 1
  - "Is the customer going to leave the service?" 0 or 1
- Examples of techniques for predicting labels: logistic regression, k-nearest neighbours, naive Bayes, discriminant analysis, classification trees, support vector machines, etc.
- We will not cover classification problems in this course, but you can study this in the Data Mining course (QBUS6810) and/or Machine Learning course (QBUS6850)

Cross-sectional data

- Cross-sectional data are values observed "at one point" in time.
  - Starting salary and WAM for graduates in 2015
  - For the used Commodores, data were collected on each car's age, condition, odometer reading and on whether the seller is an individual or a dealer.
  - Annual return on Fortune 500 company stocks in 2014.
  - Votes for or against Labor party in 2016 Federal election.

Example: car emissions

Model	Engine (litres)	City (mpg)	Highway (mpg)	Carbon (tons CO2 per year)
Chevrolet Aveo	1.6	25	34	6.6
Chevrolet Aveo 5	1.6	25	34	6.6
Honda Civic	1.8	25	36	6.3
Honda Civic Hybrid	1.3	40	45	4.4
Honda Fit	1.5	27	33	6.1
Honda Fit	1.5	28	35	5.9
Hyundai Accent	1.6	26	35	6.3
Kia Rio	1.6	26	35	6.1
Nissan Versa	1.8	27	33	6.3
Nissan Versa	1.8	24	32	6.8
Pontiac G3 Wave	1.6	25	34	6.6
Pontiac G3 Wave 5	1.6	25	34	6.6
Pontiac Vibe	1.8	26	31	6.6
Saturn Astra 2DR Hatchback	1.8	24	30	6.8
Saturn Astra 4DR Hatchback	1.8	24	30	6.8
Scion xD	1.8	26	32	6.6
Toyota Corolla	1.8	27	35	6.1
Toyota Matrix	1.8	25	31	6.6
Toyota Prius	1.5	48	45	4.0
Toyota Yaris	1.5	29	35	5.9

Cross sectional predictions

- Predict the carbon footprint (tons of CO2 per year) for other similar vehicles that are not included in the above table.
- First estimate the effects of the predictors.
- Then, forecast its carbon footprint.

Time series data

- A time series is a time stamped sequence of observations on a variable.
  - Weekly unit sales of a product.
  - Unemployment rate in Australia each quarter.
  - Daily production levels of a product.
  - Average annual temperature in Sydney.
  - Monthly water level in Warragamba Dam.
  - 5 minute prices for CBA stock on the ASX.
- Mathematically we denote a time series as, finite or infinite in time,

$$Y_0, Y_1, Y_2, \ldots, Y_t, \ldots, Y_n, Y_{n+1}, \ldots$$

What type of generalisation does your data support?

- For purposes of this course, we will be dealing with observational data and will only be interested in predicting a numerical variable of interest, not establishing causality, which is the job of econometrics or economics
- Traditional data are collected in certain design according to the purpose of research.
- Nowadays data are collected without any specific purpose pre-designed. It is up to data scientists to decide what to do with it

What is Big Data? Is this a "Big Data" course?

- No commonly accepted definition. Some ways to think about "Big Data":
  - An ever increasing amount of data that exceeds the capacity of current computational resources (memory, storage, processing, etc).
  - More data than you can deal with using a single computer.
  - A cultural phenomenon: data is an increasing part of our lives due to technological advances.
  - The Vs: volume, variety, velocity, etc.
  - A very large dataset, meaning lots of observations (large N) and/or explanatory variables/features (large p).

Themes

- An effective predictive model incorporates intuition and deep knowledge of the problem context (domain expertise).
- We must be careful not to overt by using excessively complex models that pick up noise in the training sample instead of underlying predictive patterns: Occam's Razor
- All models are wrong. No single model or method will always be best. We should consider a wide variety of techniques.
- Combining predictions from different models often works best than any single model in isolation. At first glance, this may seem like a form of overfitting, but that turns out not to be the case.

Introduction to forecasting

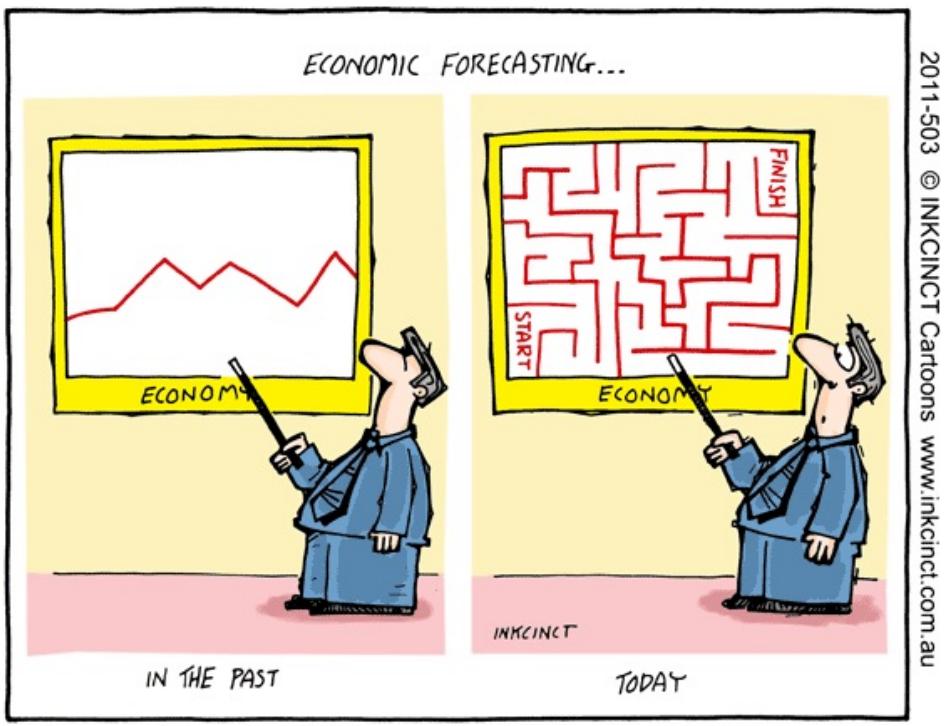
- A forecast is a prediction of what might happen in the future.
- Forecasting is a collection of methods for generating forecasts.
- "*The future aren't what it used to be!*" – Yogi Berra

Importance of forecasting

- Governments need to forecast unemployment, interest rates, expected revenues from income taxes to formulate policies.
- Companies need to forecast demand, sales, consumer preferences in strategic planning.
- Banks/investors/financial analysts need to forecast financial returns, risk or volatility, market timing.
- University administrators need to forecast enrollments to plan for facilities and for faculty recruitment
- Retail stores need to forecast demand to control inventory levels, hire employees and provide training
- Sports organisations need to project sports performance, crowd figures, club gear sales, revenues, etc. in the coming season.

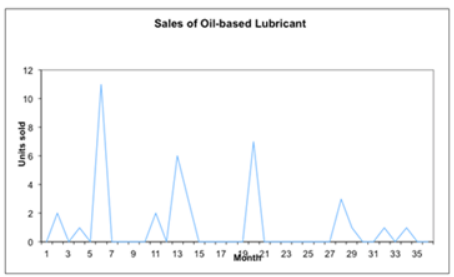
Forecasting

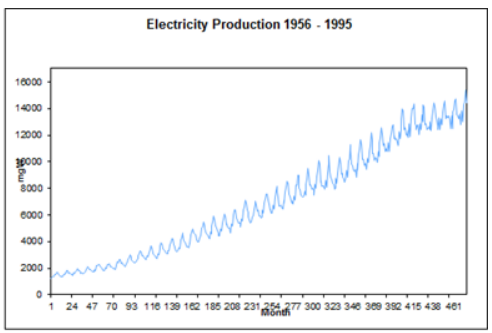
- Predicting future events and conditions.
- The act of making such predictions is called forecasting.
- Forecasting influences business and economic decision making, planning, policy setting, etc.

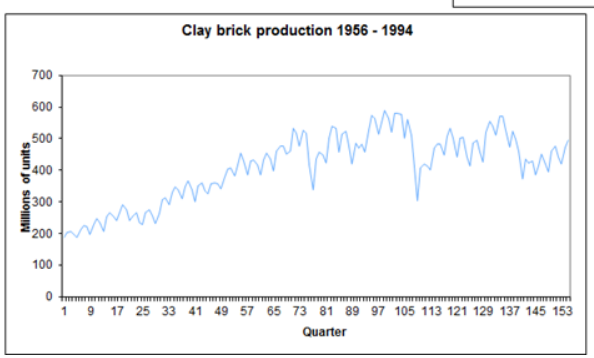


Time series plots and forecasting

- Top left plot is monthly sales of a lubricant. It is challenging to forecast because...
- Top right plot is monthly electricity production. It is easier to forecast because..
- Bottom plot is quarterly sales of clay bricks in Australia. It is challenging to forecast because...



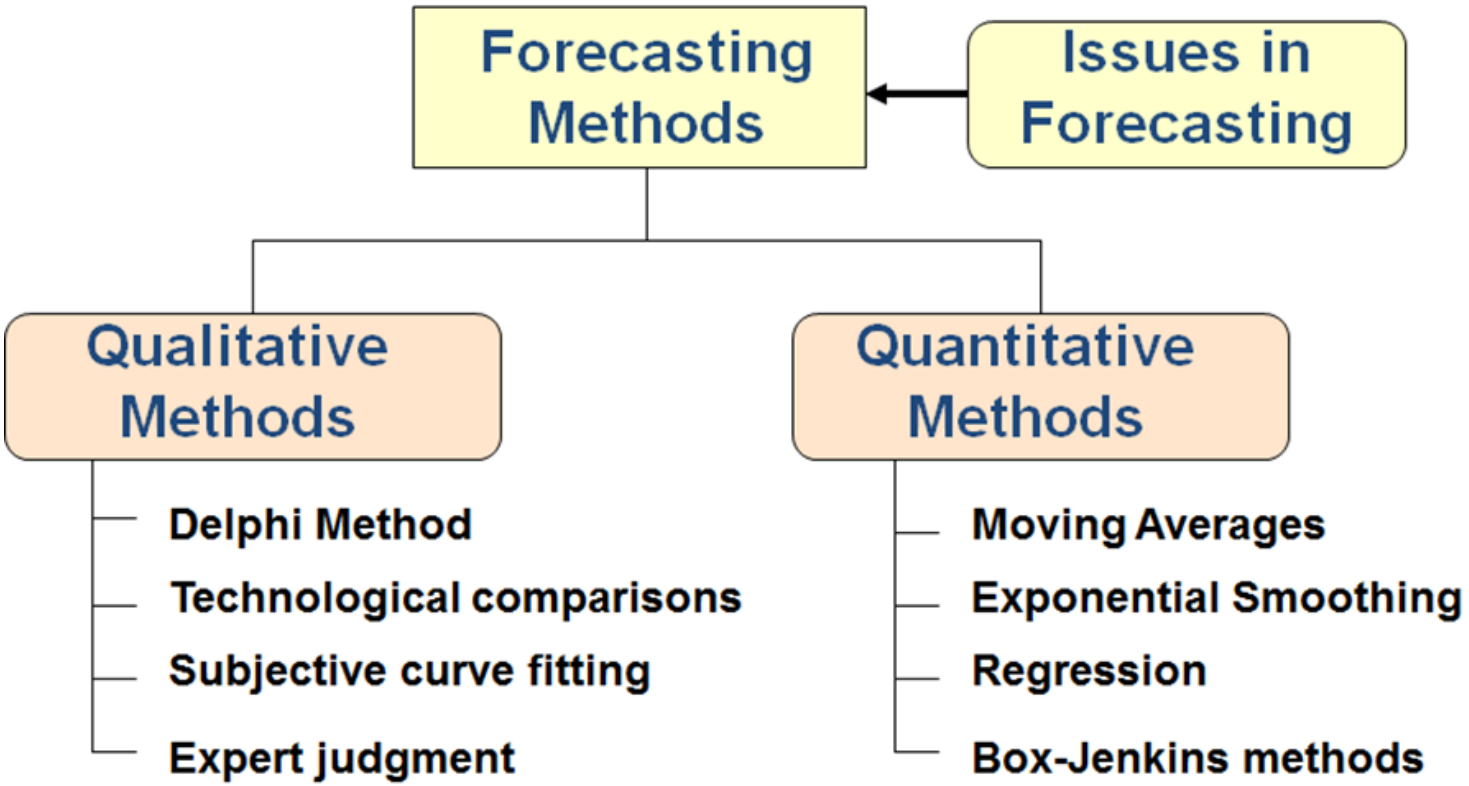




Types of forecasting

- Qualitative (judgemental) forecasting.
- Quantitative (data based) forecasting.
- Prediction markets.
- Most common approach in business: **judgementally adjusted statistical forecasting**

Types of forecasting

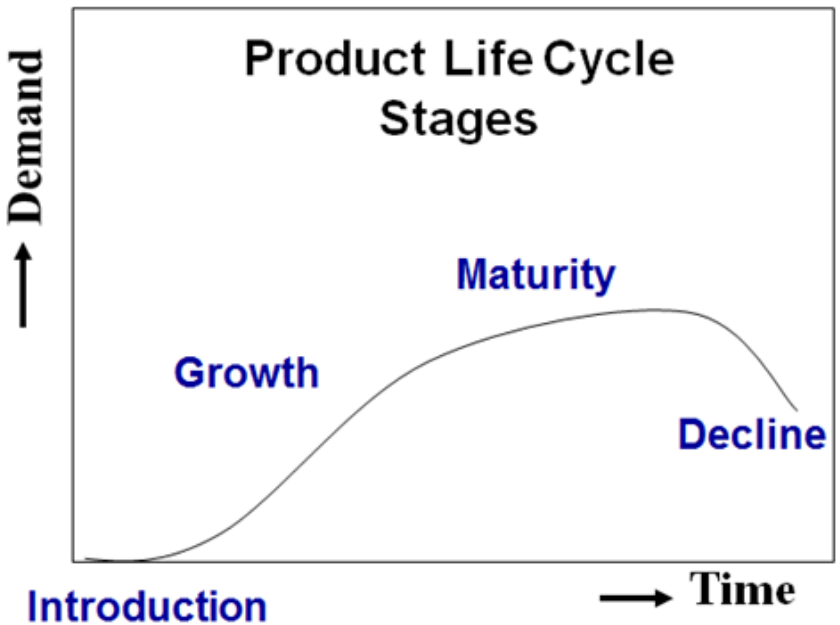


Judgemental forecasting

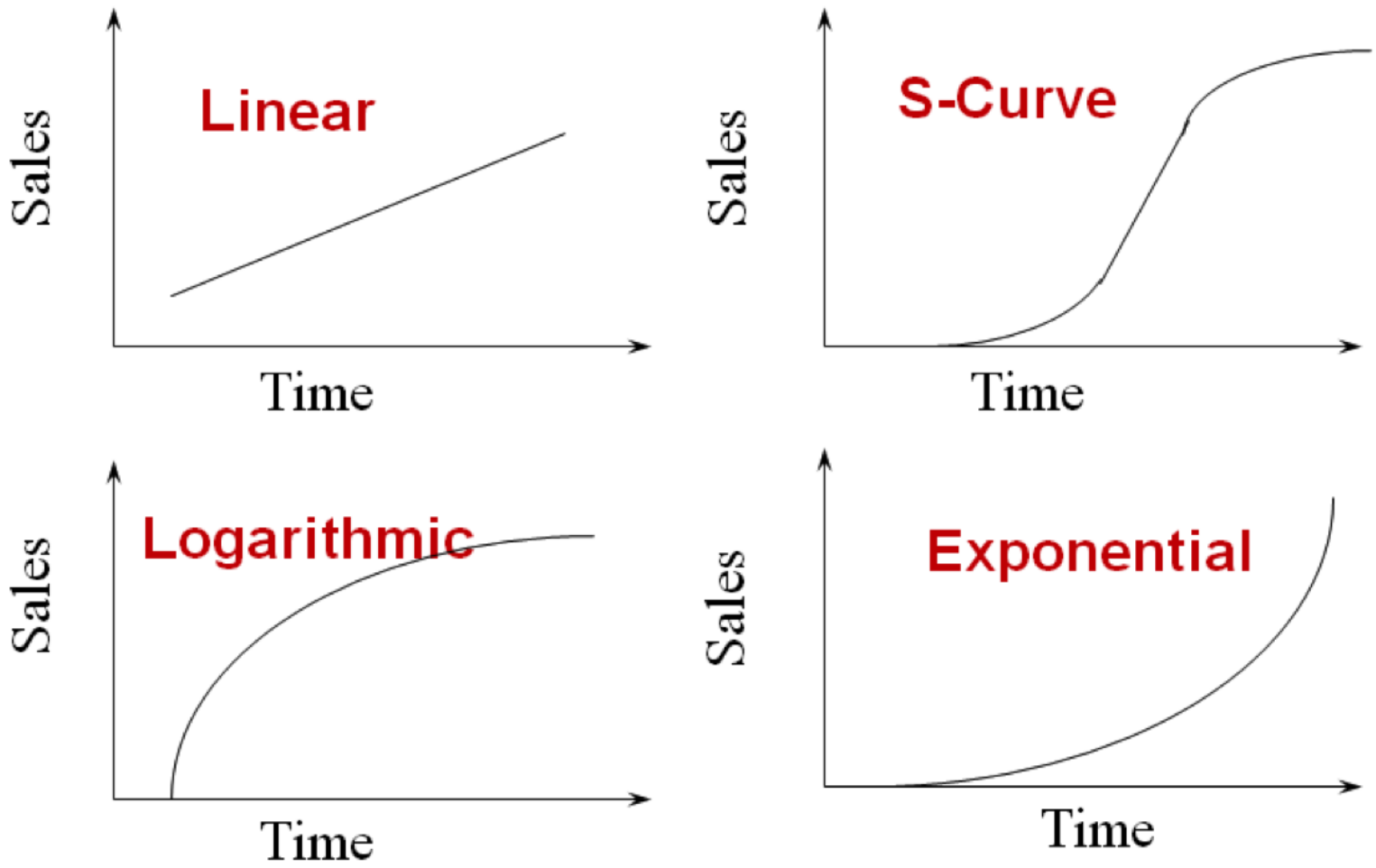
- Expert opinion (subjective)
- Subjective curve fitting
- Delphi method
  - Invented in 1950s by Helmer and Dalkey
  - Assumption: forecasts from a group is more accurate than those from individuals
  - Stages: forming panel, setting tasks, initial expert views, feedback to experts, aggregating expert views for forecasting
- Subjectively extending previous patterns into future

Subjective curve fitting (e.g. product life cycle)

- The life cycle of a product typically involves four stages: introduction, growth, maturity, and decline.
- Note that forecasts in different stages of the life cycle require different approaches.



Typical sales forecasts of products during the growth stage



The Delphi method

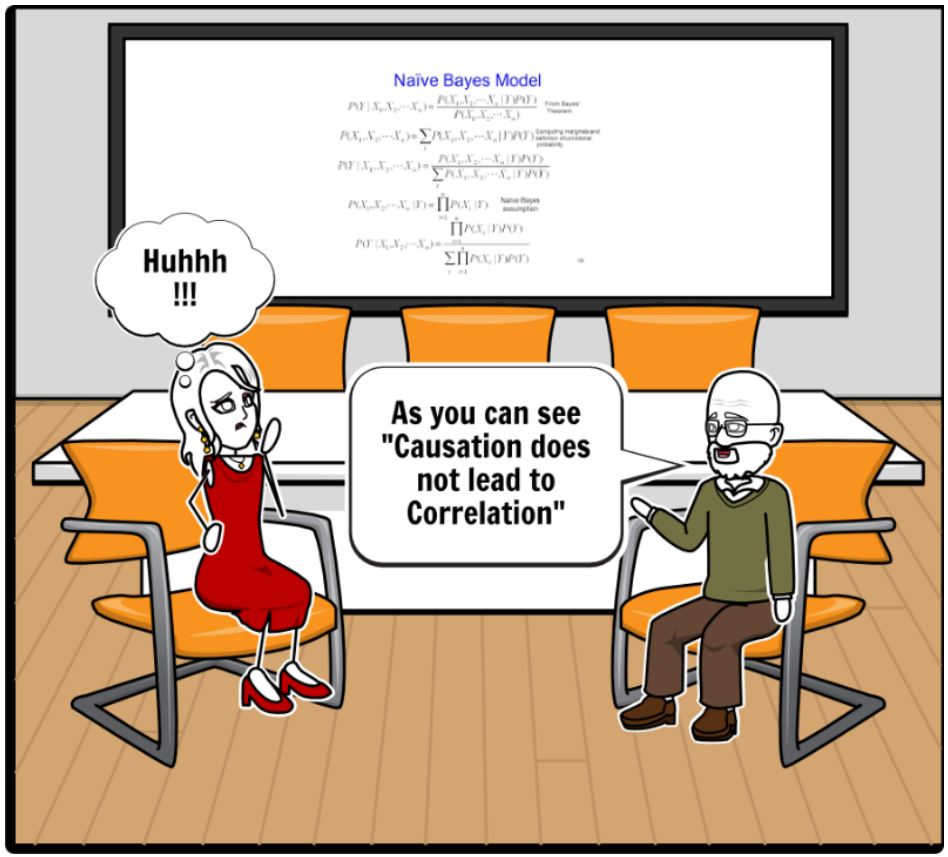
1. Panel of experts BUT never meet as a group.
2. Use questionnaire to obtain forecasts from all participants
3. Summarize and re-distribute results to all participants PLUS appropriate new questions
4. Summarize again, refining forecasts and conditions, and develop new questions.
5. Repeat Step 4 if necessary. Distribute the final results to all participants.

Prediction markets

- Prediction markets seek to tap into the "wisdom of the crowds". In prediction markets, participants can effectively make bets on particular outcomes, such as the sales of a new product or whether a project will be finished on time. The market price of an event should then reflect the consensus probability that its event will occur. There is still debate on whether this holds in theory and practice.
- They are used by Google, HP, and other big companies.
- The prediction markets for politics https://www.predictit.org/, e.g., https://www.predictit.org/Contract/5534/Will-Facebook's-Mark-Zuckerberg-run-for-president-in-2020#data

Quantitative Forecasting

- Formal econometric or statistical forecasting methods.
  - Project previous patterns into future using a statistical model.
  - Time series modelling.
  - Regression (causal) modelling.
- Less formal (adhoc) or intuitive approaches.
- Naïve forecasting model.



Naïve (random walk) forecasts

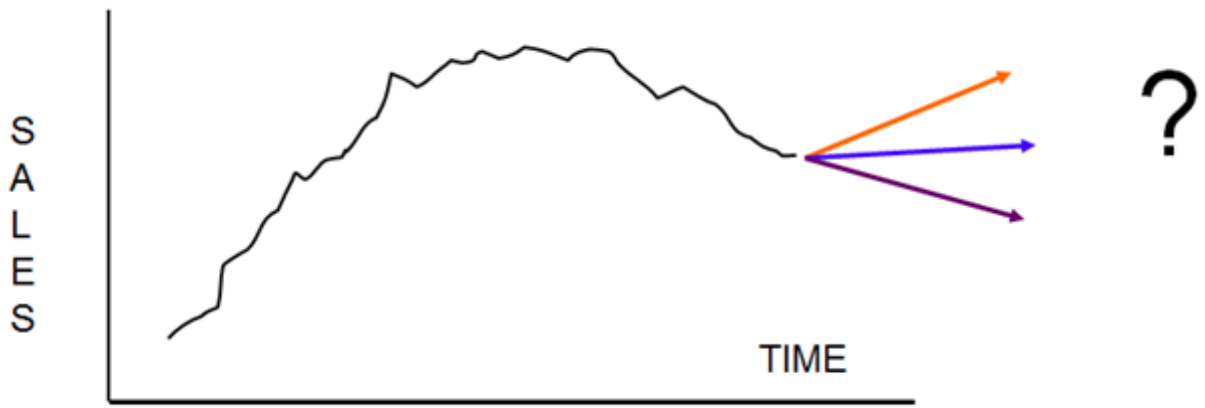
The forecast for any period equals the previous period's actual value.

Time series forecasts

- Project previous patterns into future using a formal statistical model e.g.

$$\text{Sales}_t = f(\text{Sales}_{t-1}, \text{Sales}_{t-2}, ..., \text{Sales}_{t-24}) + \epsilon_t$$

- Only concerned with forecasting, not reasons why the variable changes

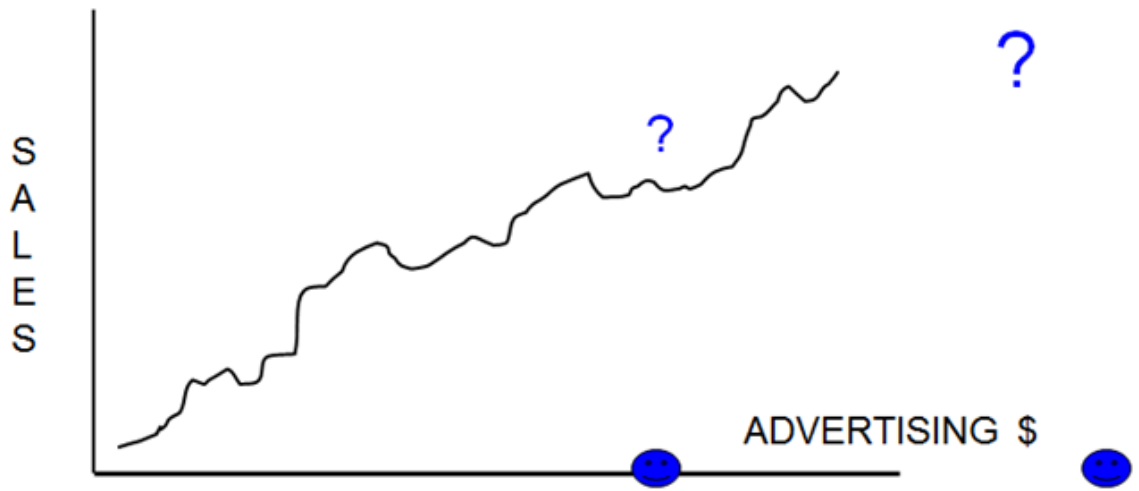


Regression forecasts

- Use a formal statistical regression model

$$\text{Price}_t = f(\text{Season}_t; \text{Demand}_t; \text{GeneratorVendor}_t) + \epsilon_t$$

- Can assess other quantities related to price changes
- Can do scenario forecasting



Time series regression

- Models can combine time series and regression components:

$$\begin{aligned}\text{Price}_t =& f(\text{Season}_t, \text{Demand}_t, \text{Price}_{t-1}, \ldots, \text{Price}_{t-12}) + \epsilon_t \\ &+ g(\epsilon_{t-1}, \epsilon_{t-2}, \epsilon_{t-3})\end{aligned}$$

The process of forecasting

- Six principles of Forecasting (Armstrong, 2001):
  - Problem formulation
  - Obtain information.
  - Choose method.
  - Forecast.
  - Assess forecasts.
  - Implement decision making.
- 139 sub-principles of forecasting!

The process of forecasting

- FPP2 (https://otexts.com/fpp2/basic-steps.html). Five basic steps in forecasting:
  - Problem definition
  - Gathering information
  - Preliminary data analysis
  - Choose and test models.
  - Using and evaluating a forecasting model (forecast, assess forecasts, and implement decision making).
- I would add: define performance metric after gathering information. You may also want to build a data product at the end.

Problem formulation/definition

- Formulation
  - What exactly needs forecasting?
  - Can the variable of interest even be forecasted?
  - What information is available?
- Principles
  - Use experts knowledge/previous studies to examine if forecasting is considered possible.
  - Use theory to guide the search for possible explanatory factors.
  - Communicate with all involved in data collection, decision making, etc. to properly structure problem definition.

Problem formulation/definition

- Defintion
  - How will forecasts be used?
  - What decisions might result?
  - Objectives?
- More principles
  - Decision makers should have agreed upon actions based on certain forecast results.
  - Decision makers should agree on method to be used.

Case 2 (https://otexts.com/fpp2/case-studies.html)

- The Australian federal government needed to forecast the annual budget for the Pharmaceutical Benet Scheme (PBS).
- The PBS provides a subsidy for many pharmaceutical products sold in Australia.
- The total expenditure was around $7 billion in 2009 and had been underestimated by nearly $1 billion in the each of the two years.
- In order to forecast the total expenditure, it is necessary to forecast hundreds of groups.
- So we needed to find a forecasting method that allowed for trend and seasonality if it was present, and that was robust to sudden changes in the underlying patterns.
- How might we go about defining the problem here?

Where to find time series data on the web? Australian Data

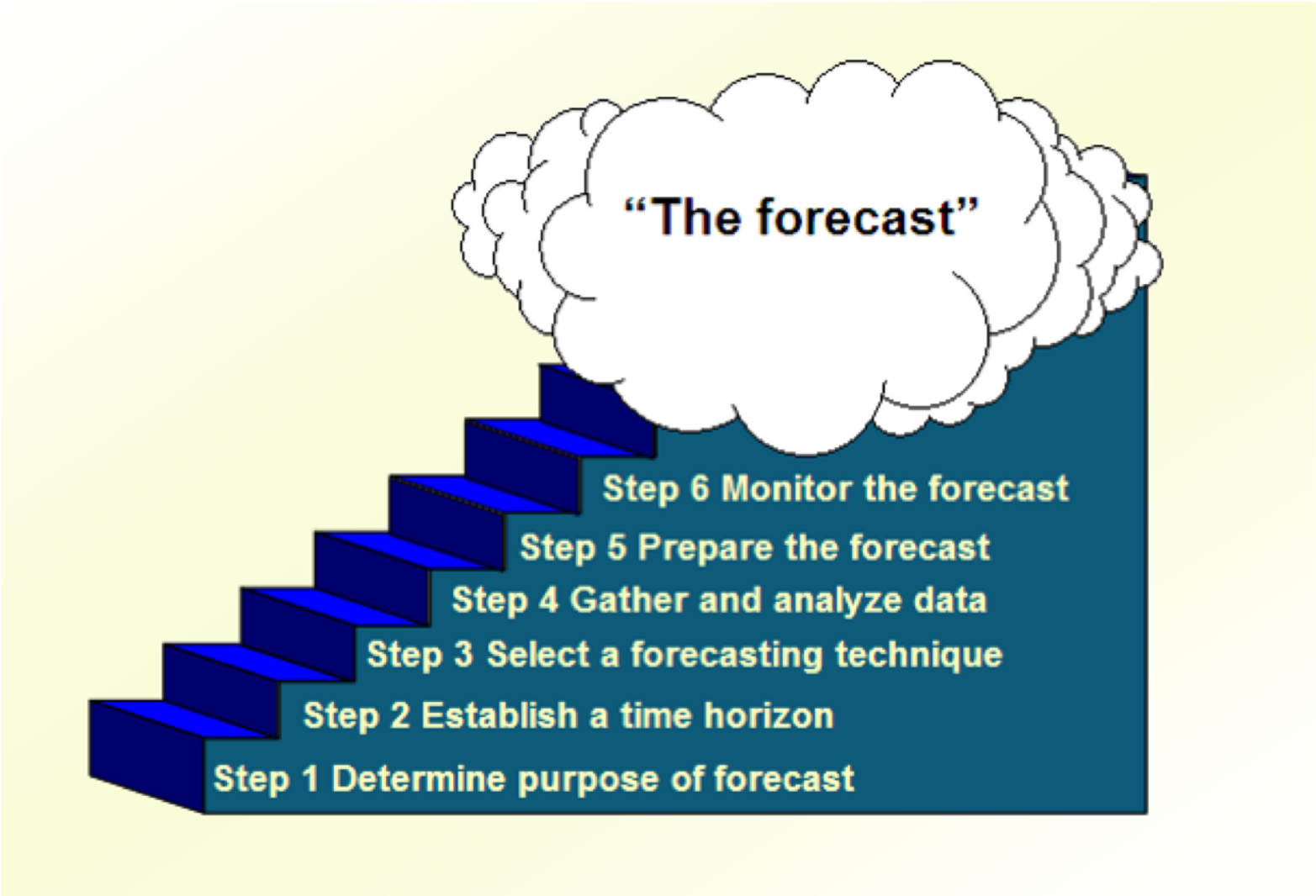
- Data Libraries
- OZDasl - Australian Data and Story Library
- ANU Social Science Data Archives
- University of Sydney Library Databases Collection
- Original Data Sources
- Reserve Bank of Australia
- Australian Bureau of Statistics
- Penn World Tables ? Australia
- Time Series Data Library (Hyndman)
- Datastream International

Where to find time series data on the web? US Data

- EconData
- Data and Story Library — CMU
- Bureau of Economic Analysis
- Economagic - times series data
- FRED - Federal Reserve Economic Data
- White House - Economic Statistics Brieng Room
- NBER - National Bureau of Economic Research
- NBER - Marriage and Divorce Data
- ICensus - Statistical Abstracts
- Census Data
- ICPSR - Interuniversity Consortium for Social and Political Research
- Panel Survey of Income Dynamics (PSID)
- Bureau of Labor Statistics
- Survey of Income and Program Participation (SIPP)
- National Center for Health Statistics
- Statistics in Sport (American Statistical Association)

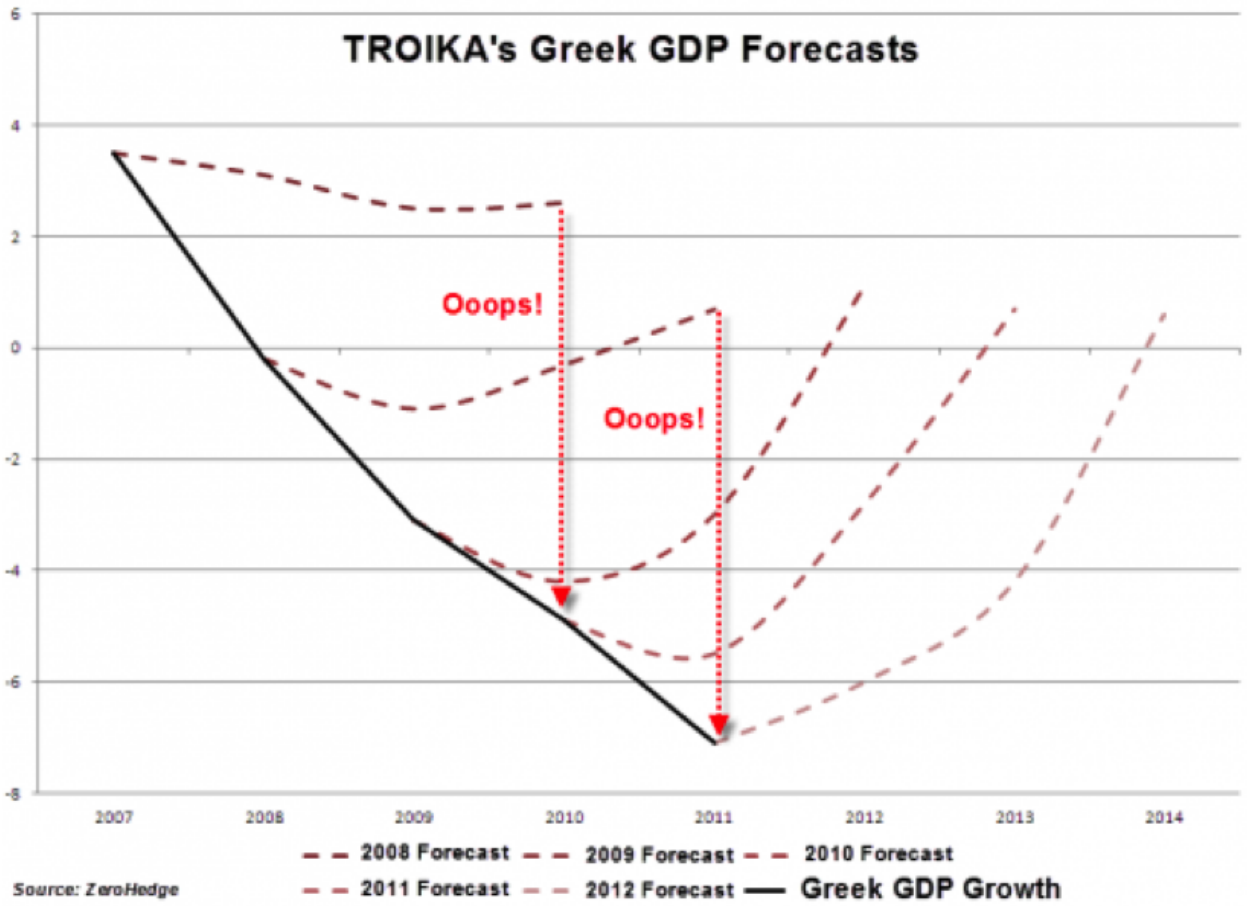
Elements of a Good Forecast

Steps in the Forecasting Process



Some more principles

- *It is difficult to make predictions, especially about the future* – Unknown Danish author



Source: Zerohedge, via Felix Salmon's Twitter