

# Predictive Analytics

Ch1. Getting started

Prof. Dr. Benjamin Buchwitz

Wir geben Impulse

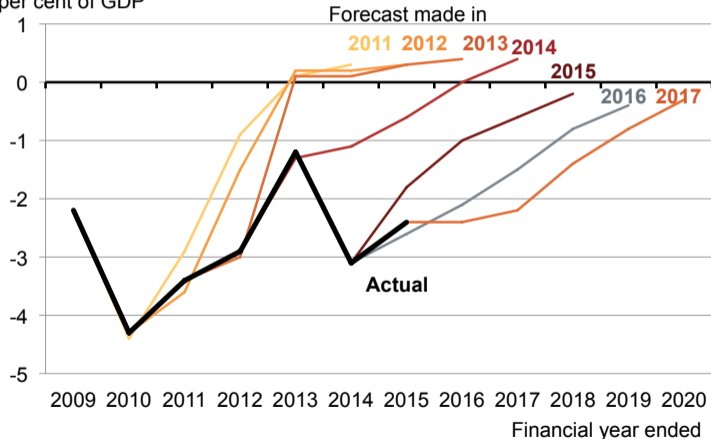


- 1 What can we forecast?
- 2 Time series data and random futures
- 3 Some case studies
- 4 Assignment 1

## Commonwealth plans to drift back to surplus show the triumph of experience over hope

GRATTAN  
Institute

Actual and forecast Commonwealth underlying cash balance  
per cent of GDP



# What can we forecast?



# What can we forecast?



What can we forecast?



What can we forecast?



# What can we forecast?





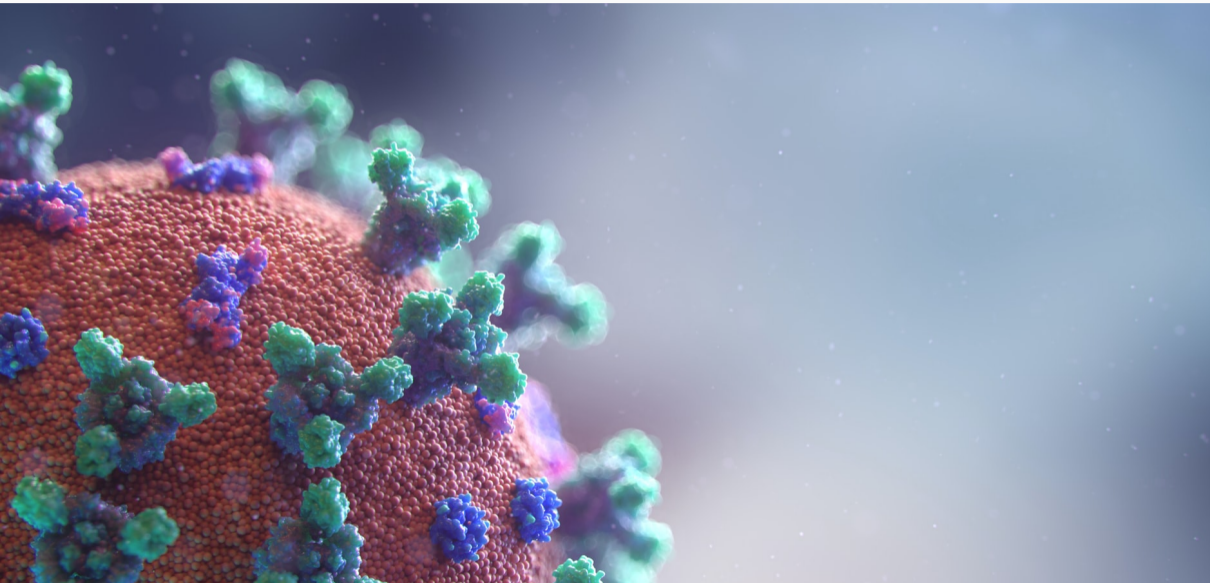
# What can we forecast?



# What can we forecast?



# What can we forecast?



# What can we forecast?



## Which is easiest to forecast?

- daily electricity demand in 3 days time
- time of sunrise this day next year
- Google stock price tomorrow
- Google stock price in 6 months time
- maximum temperature tomorrow
- exchange rate of \$US/AUS next week
- total sales of drugs in Australian pharmacies next month
- timing of next Halley's comet appearance

## Which is easiest to forecast?

- daily electricity demand in 3 days time
- time of sunrise this day next year
- Google stock price tomorrow
- Google stock price in 6 months time
- maximum temperature tomorrow
- exchange rate of \$US/AUS next week
- total sales of drugs in Australian pharmacies next month
- timing of next Halley's comet appearance

[pollev.com/robjhyndman](https://pollev.com/robjhyndman)

## Which is easiest to forecast?

- 1 time of sunrise this day next year
- 2 timing of next Halley's comet appearance
- 3 maximum temperature tomorrow
- 4 daily electricity demand in 3 days time
- 5 total sales of drugs in Australian pharmacies next month
- 6 Google stock price tomorrow
- 7 exchange rate of \$US/AUS next week
- 8 Google stock price in 6 months time

## Which is easiest to forecast?

- 1 time of sunrise this day next year
- 2 timing of next Halley's comet appearance
- 3 maximum temperature tomorrow
- 4 daily electricity demand in 3 days time
- 5 total sales of drugs in Australian pharmacies next month
- 6 Google stock price tomorrow
- 7 exchange rate of \$US/AUS next week
- 8 Google stock price in 6 months time

- how do we measure "easiest"?
- what makes something easy/difficult to forecast?



Something is easier to forecast if:

- 1 we have a good understanding of the factors that contribute to it
- 2 there is lots of data available;
- 3 the future is somewhat similar to the past
- 4 the forecasts cannot affect the thing we are trying to forecast.

- 1 What can we forecast?
- 2 Time series data and random futures
- 3 Some case studies
- 4 Assignment 1

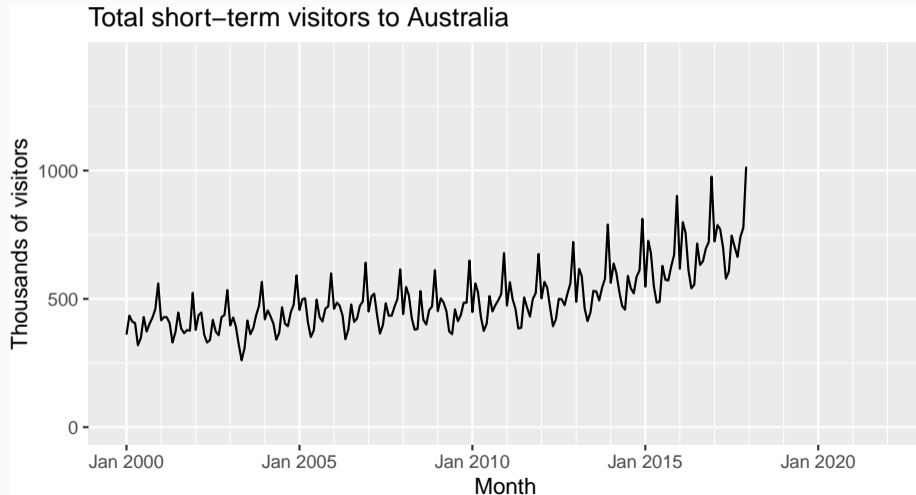
- Four-yearly Olympic winning times
- Annual Google profits
- Quarterly Australian beer production
- Monthly rainfall
- Weekly retail sales
- Daily IBM stock prices
- Hourly electricity demand
- 5-minute freeway traffic counts
- Time-stamped stock transaction data

## Random futures

A forecast is an estimate of the probabilities of possible futures.

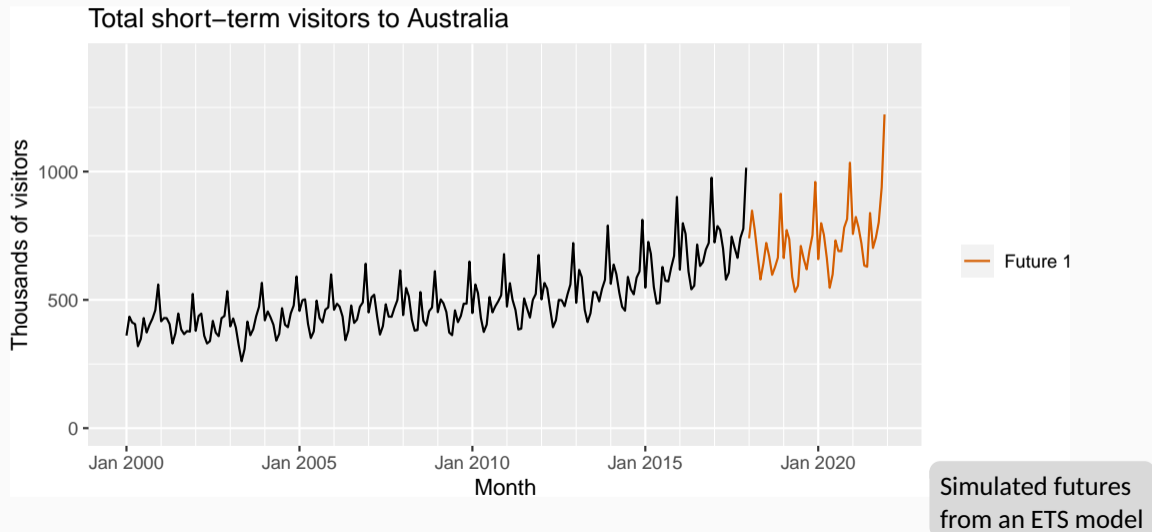
# Random futures

A forecast is an estimate of the probabilities of possible futures.



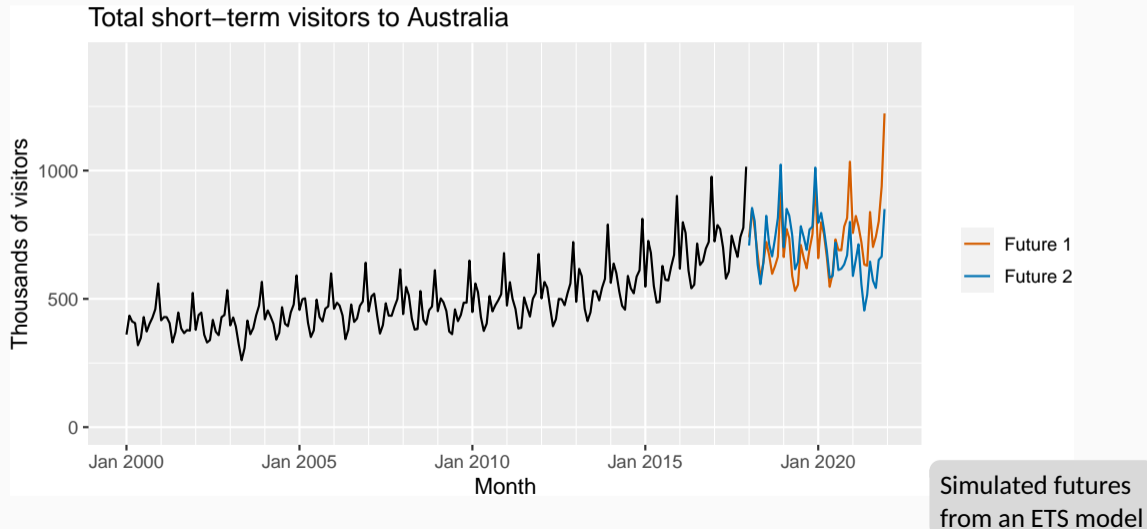
# Random futures

A forecast is an estimate of the probabilities of possible futures.



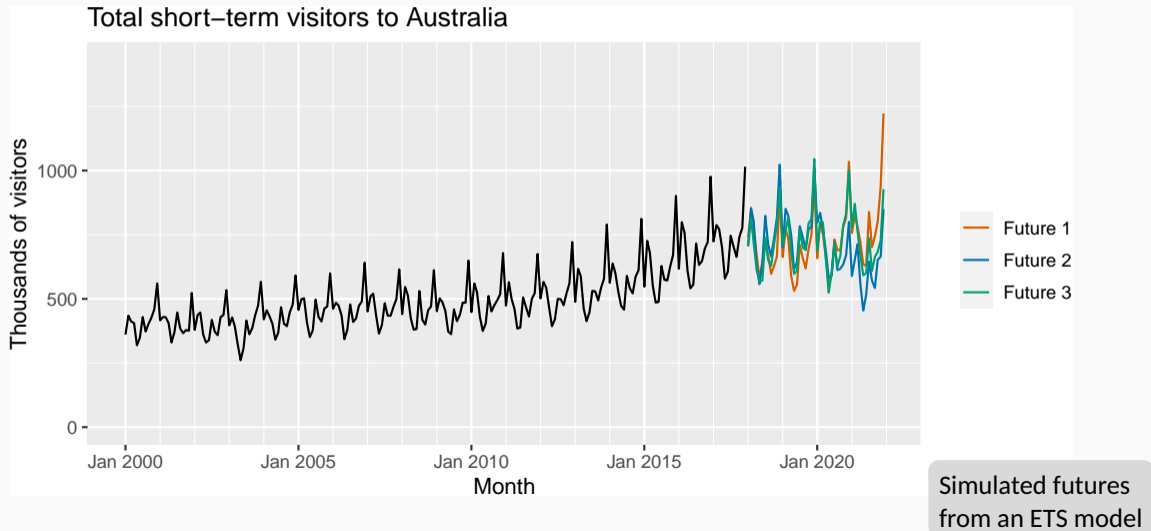
# Random futures

A forecast is an estimate of the probabilities of possible futures.



# Random futures

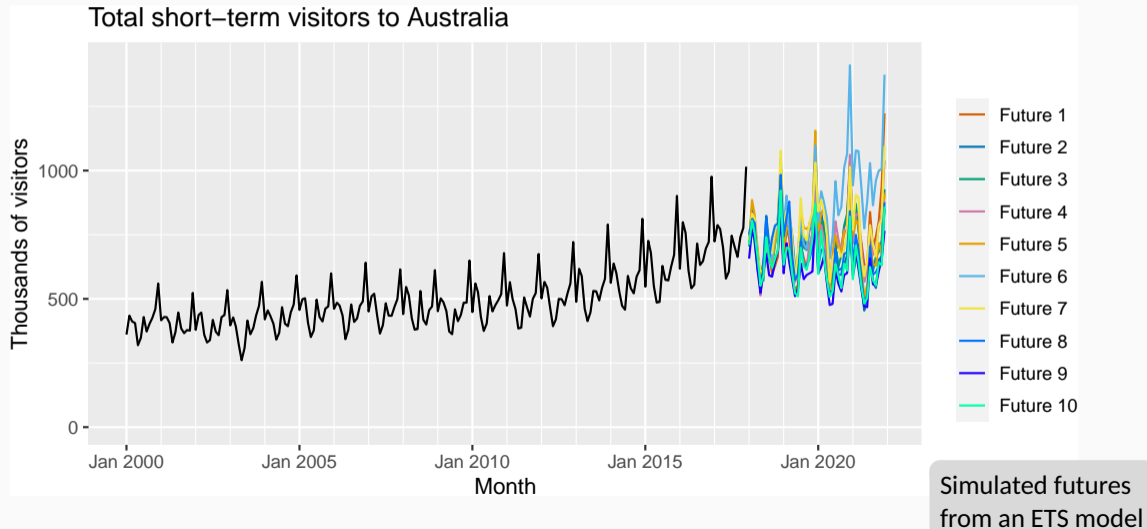
A forecast is an estimate of the probabilities of possible futures.





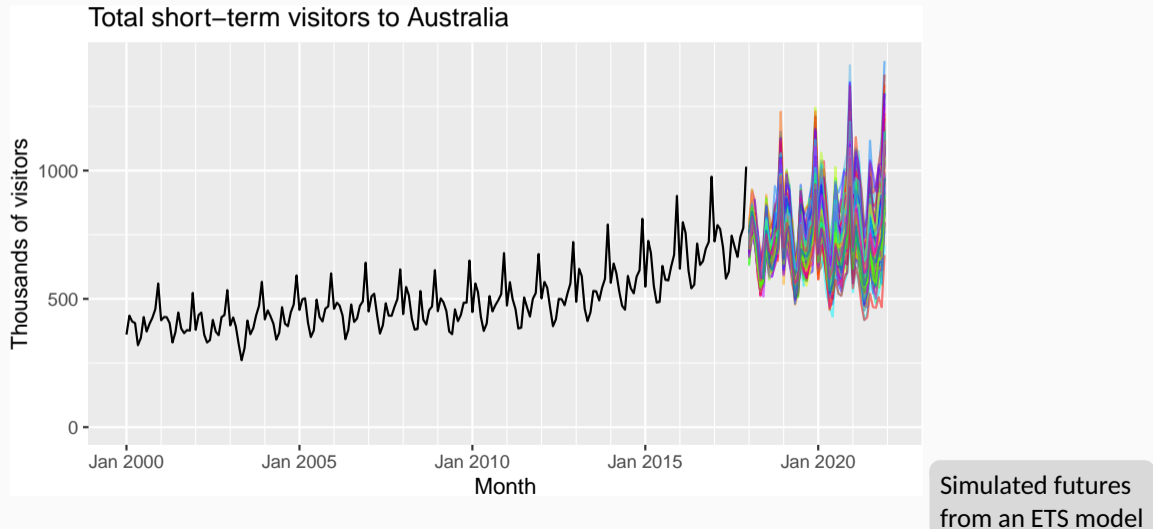
# Random futures

A forecast is an estimate of the probabilities of possible futures.



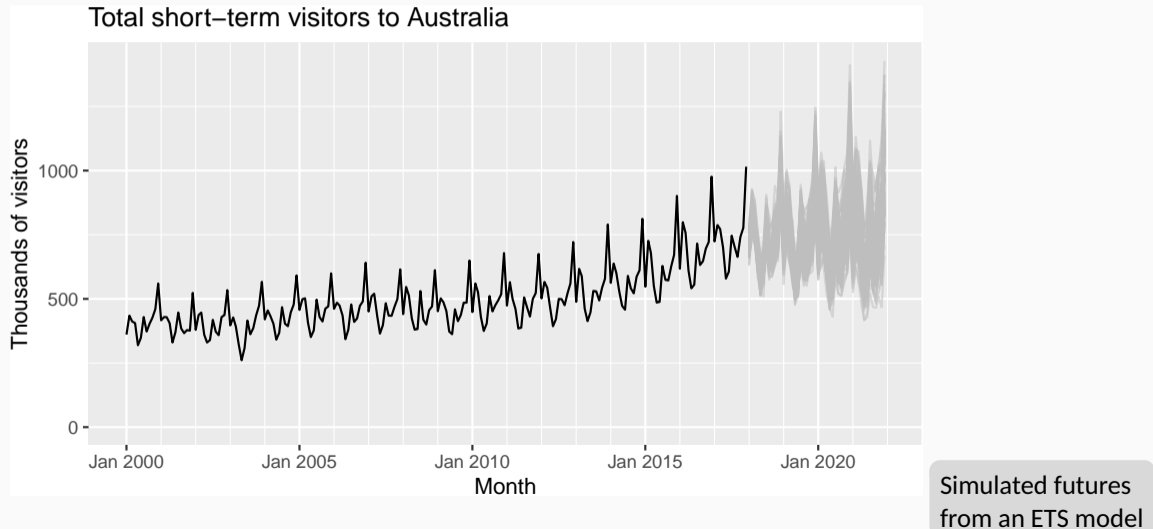
# Random futures

A forecast is an estimate of the probabilities of possible futures.



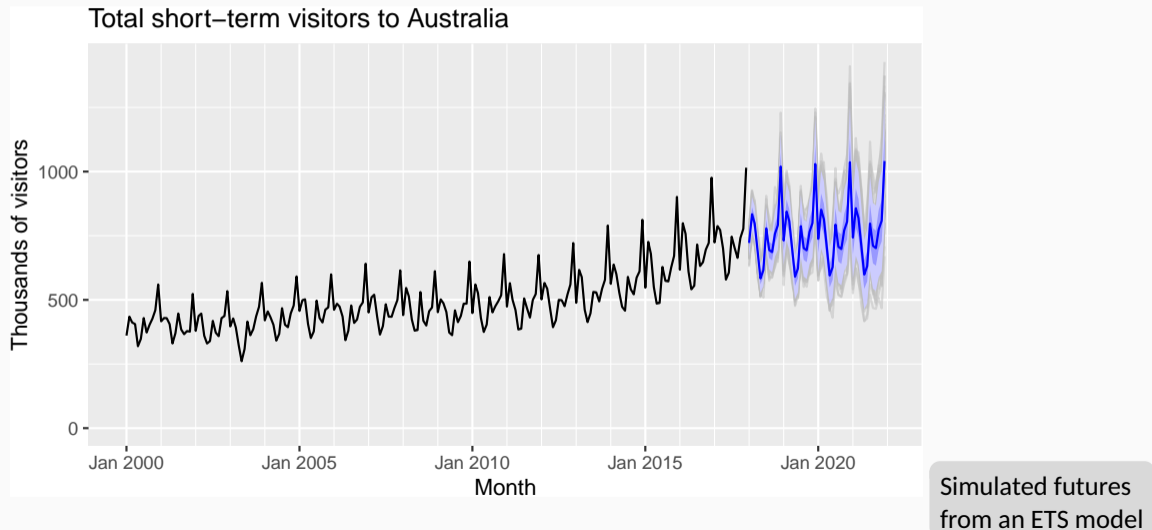
# Random futures

A forecast is an estimate of the probabilities of possible futures.



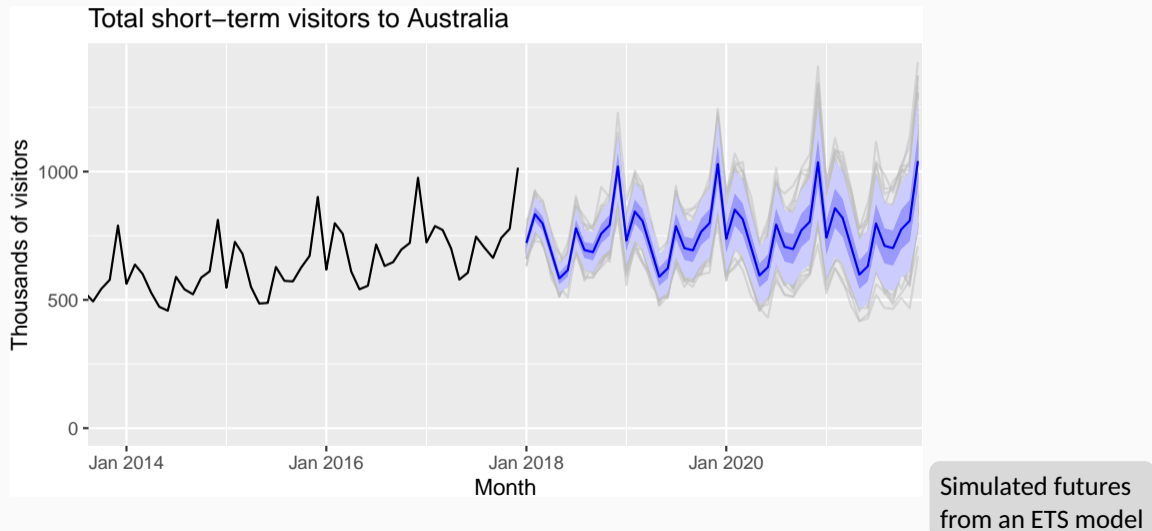
# Random futures

A forecast is an estimate of the probabilities of possible futures.



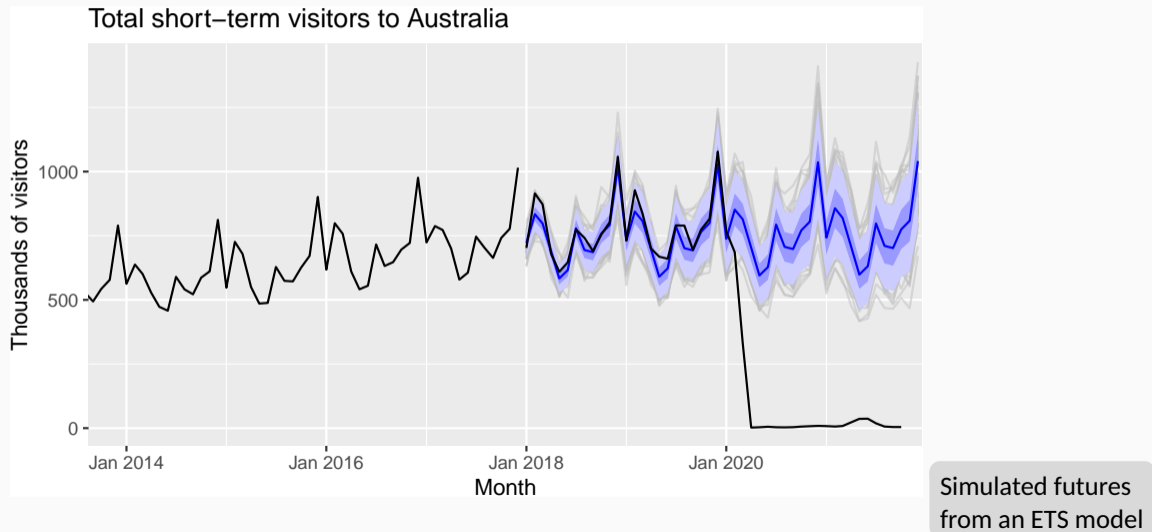
# Random futures

A forecast is an estimate of the probabilities of possible futures.



# Random futures

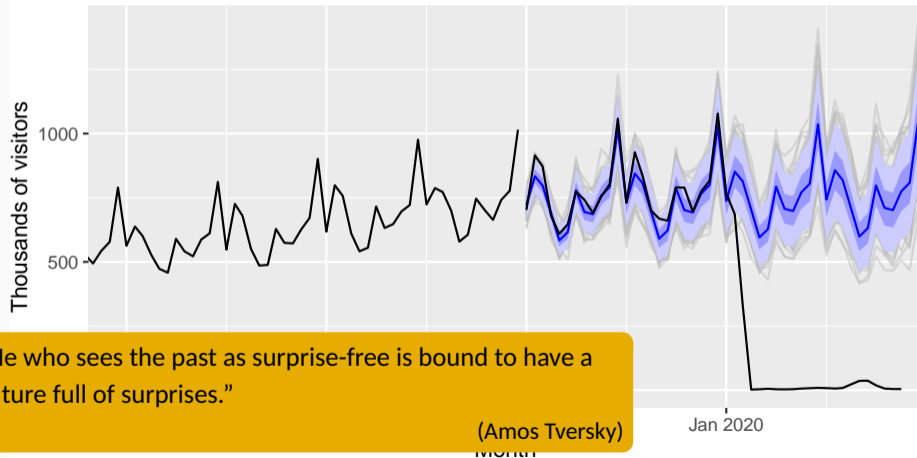
A forecast is an estimate of the probabilities of possible futures.



# Random futures

A forecast is an estimate of the probabilities of possible futures.

Total short-term visitors to Australia



“He who sees the past as surprise-free is bound to have a future full of surprises.”

(Amos Tversky)

Simulated futures  
from an ETS model

- Thing to be forecast: a random variable,  $y_t$ .
- Forecast distribution: If  $\mathcal{I}$  is all observations, then  $y_t|\mathcal{I}$  means “the random variable  $y_t$  given what we know in  $\mathcal{I}$ .”
- The “point forecast” is the mean (or median) of  $y_t|\mathcal{I}$
- The “forecast variance” is  $\text{var}[y_t|\mathcal{I}]$
- A prediction interval or “interval forecast” is a range of values of  $y_t$  with high probability.
- With time series,  $y_{t|t-1} = y_t|\{y_1, y_2, \dots, y_{t-1}\}$ .
- $\hat{y}_{T+h|T} = E[y_{T+h}|y_1, \dots, y_T]$  (an  $h$ -step forecast taking account of all observations up to time  $T$ ).



- 1 What can we forecast?
- 2 Time series data and random futures
- 3 Some case studies
- 4 Assignment 1

## CASE STUDY 1: Paperware company

**Problem:** Want forecasts of each of hundreds of items. Series can be stationary, trended or seasonal. They currently have a large forecasting program written in-house but it doesn't seem to produce sensible forecasts. They want me to fix it.

### Additional information

- Program written in COBOL making numerical calculations limited. It is not possible to do any optimisation.
- Their programmer has little experience in numerical computing.
- They employ no statisticians and want the program to produce forecasts automatically.



### Methods currently used

- A** 12 month average
- C** 6 month average
- E** straight line regression over last 12 months
- G** straight line regression over last 6 months
- H** average slope between last year's and this year's values. (Equivalent to differencing at lag 12 and taking mean.)
- I** Same as H except over 6 months.
- K** I couldn't understand the explanation.

## CASE STUDY 2: PBS



### The Pharmaceutical Benefits Scheme (PBS) is the Australian government drugs subsidy scheme.

- Many drugs bought from pharmacies are subsidised to allow more equitable access to modern drugs.
- The cost to government is determined by the number and types of drugs purchased. Currently nearly 1% of GDP.
- The total cost is budgeted based on forecasts of drug usage.



# ABC News Online

AUSTRALIAN BROADCASTING CORPORATION



## NewsRadio

Streaming audio news

LISTEN: [WMP](#) | [Real](#)

Select a Topic from the list below

- [Top Stories](#)
- [Just In](#)
- [World](#)
- [Asia-Pacific](#)
- [Business](#)
- [Sport](#)
- [Arts](#)
- [Sci Tech](#)
- [Indigenous](#)
- [Weather](#)
- [Rural](#)
- [Local News](#)
- [Broadband](#)

Click "Refresh" or "Reload" on your browser for the latest edition.

This Bulletin: **Wed, May 30 2001 6:22 PM AEST**

## POLITICS

### Opp demands drug price restriction after PBS budget blow-out

The Federal Opposition has called for tighter controls on drug prices after the Pharmaceutical Benefits Scheme (PBS) budget blew out by almost \$800 million.

The money was spent on two new drugs including the controversial anti-smoking aid Zyban, which dropped in price from \$220 to \$22 after it was listed on the PBS.



**the Public Record**  
For full election coverage

### FEATURES



**Public Record**  
Federal Election 2001

[For a fresh perspective on the federal election, reach into ABC Online's campaign weblog, \*The Poll Vault\*.](#)

Audio News Online

- In 2001: \$4.5 billion budget, under-forecasted by \$800 million.
- Thousands of products. Seasonal demand.
- Subject to covert marketing, volatile products, uncontrollable expenditure.
- Although monthly data available for 10 years, data are aggregated to annual values, and only the first three years are used in estimating the forecasts.
- All forecasts being done with the FORECAST function in MS-Excel!

## CASE STUDY 3: Car fleet company

**Client:** One of Australia's largest car fleet companies

**Problem:** how to forecast resale value of vehicles? How should this affect leasing and sales policies?



## CASE STUDY 3: Car fleet company

**Client:** One of Australia's largest car fleet companies

**Problem:** how to forecast resale value of vehicles? How should this affect leasing and sales policies?

### Additional information

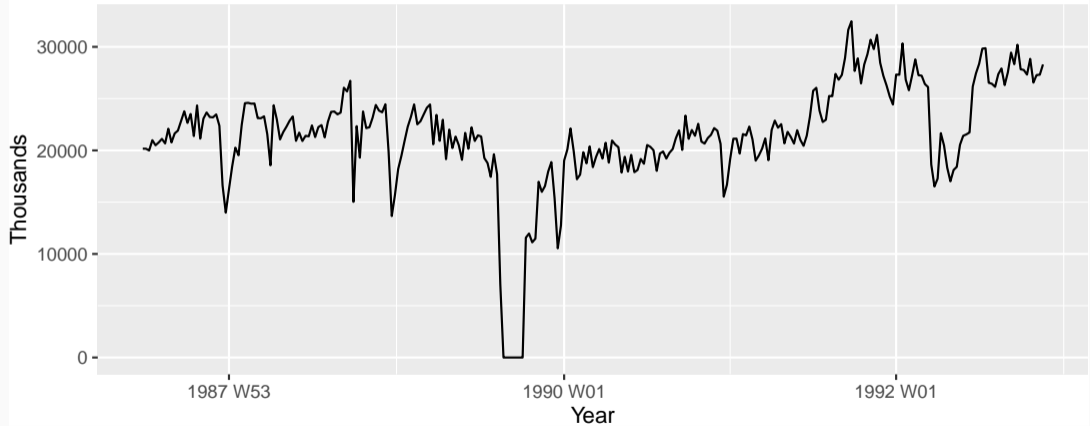
- They can provide a large amount of data on previous vehicles and their eventual resale values.
- The resale values are currently estimated by a group of specialists. They see me as a threat and do not cooperate.

## CASE STUDY 4: Airline



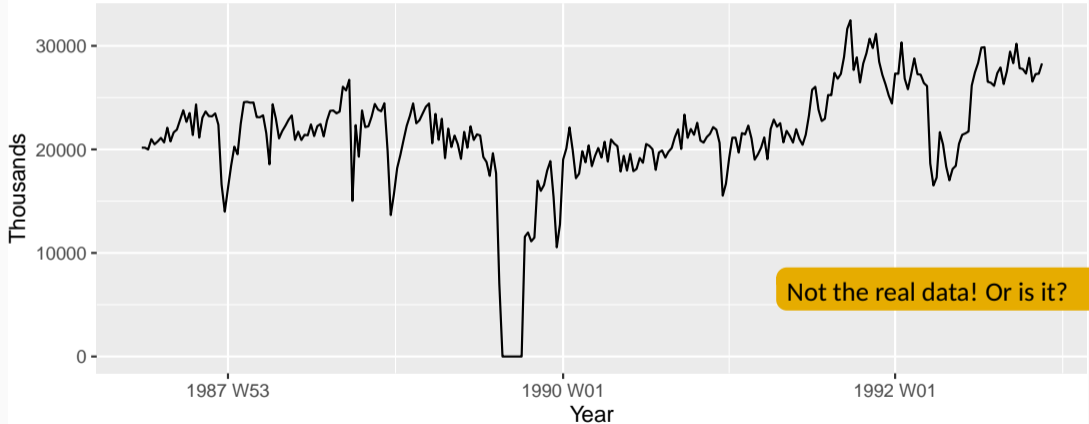
## CASE STUDY 4: Airline

Economy class passengers  
Melbourne–Sydney



## CASE STUDY 4: Airline

Economy class passengers  
Melbourne–Sydney



Not the real data! Or is it?

**Problem:** how to forecast passenger traffic on major routes?

### Additional information

- They can provide a large amount of data on previous routes.
- Traffic is affected by school holidays, special events such as the Grand Prix, advertising campaigns, competition behaviour, etc.
- They have a highly capable team of people who are able to do most of the computing.

- 1 What can we forecast?
- 2 Time series data and random futures
- 3 Some case studies
- 4 Assignment 1

## Assignment 1: forecast the following series

- 1 Google closing stock price on 15 April 2022.
- 2 Maximum temperature at Melbourne airport on 15 April 2022.
- 3 The difference in points (Collingwood minus Essendon) scored in the AFL match between Collingwood and Essendon for the Anzac Day clash. 25 April 2022.
- 4 The seasonally adjusted estimate of total employment for April 2022. ABS CAT 6202, to be released around mid May 2022.
- 5 Google closing stock price on 23 May 2022.

# Assignment 1: scoring

$Y$  = actual,  $F$  = point forecast,  $[L, U]$  = prediction interval

## Point forecasts:

$$\text{Absolute Error} = |Y - F|$$

- Rank results for all students in class
- Add ranks across all five items

## Prediction intervals:

$$\text{Interval Score} = (U - L) + 10(L - Y)_+ + 10(Y - U)_+$$

- Rank results for all students
- Add ranks across all five items