MONASH University

MONASH BUSINESS SCHOOL

ETC3550
Applied forecasting for business and economics

Ch1. Getting started

OTexts.org/fpp3/

Outline

1. What can we forecast?
2. Time series data
3. Some case studies
4. The statistical forecasting perspective
5. Assignment 1

Outline

1. What can we forecast?
2. Time series data
3. Some case studies
4. The statistical forecasting perspective
5. Assignment 1

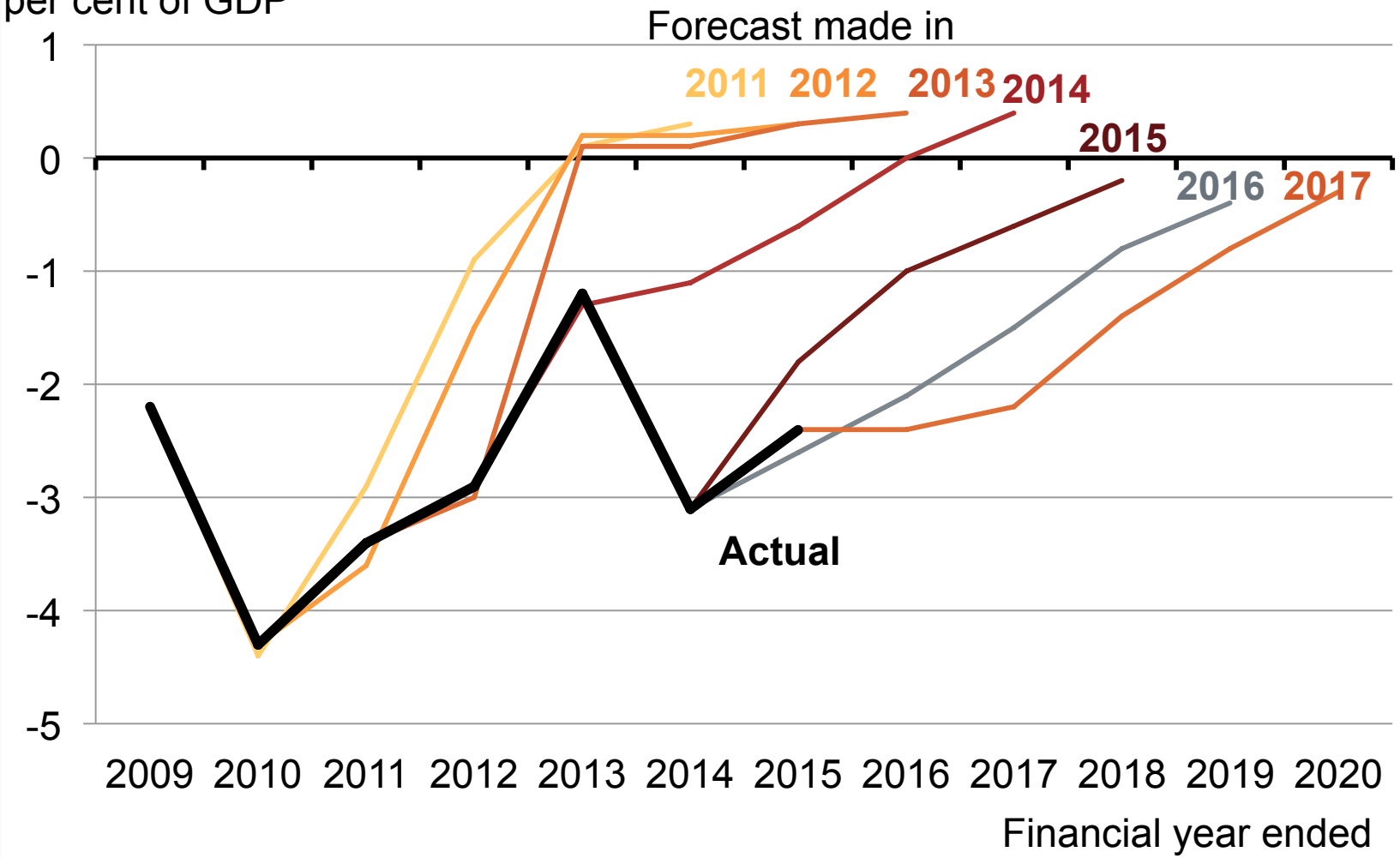
Forecasting is difficult

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



Forecasting is difficult

A Timeline of

Very Bad Future Predictions

1800

"Rail travel at high speed is not possible, because passengers, unable to breathe, would die of asphyxia."

Dr. Dionysys Larder, Professor of Natural Philosophy & Astronomy, University College London

1859

"Drill for oil? You mean drill into the ground to try and find oil? You're crazy!"

Associates of Edwin L. Drake refusing his suggestion to drill for oil in 1859 (Later that year, Drake succeeded in drilling the first oil well.)

1876

"This telephone has too many shortcomings to be seriously considered as a means of communication."

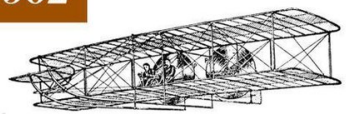
Western Union internal memo

1880

"Everyone acquainted with the subject will recognize it as a conspicuous failure."

Henry Morton, president of the Stevens Institute of Technology, on Edison's light bulb

1902



"Flight by machines heavier than air is unpractical and insignificant, if not utterly impossible."

Simon Newcomb, Canadian-American astronomer and mathematician, 18 months before the Wright Brothers' flight at Kittyhawk

1903

"The horse is here to stay, but the automobile is only a novelty, a fad."

The president of the Michigan Savings Bank, advising Henry Ford's lawyer not to invest in the Ford Motor Company

1916

"The idea that cavalry will be replaced by these iron coaches is absurd. It is little short of treasonous."

Comment of Aide-de-camp to Field Marshal Haig, at tank demonstration

1916

"The cinema is little more than a fad. It's canned drama. What audiences really want to see is flesh and blood on the stage."

Charlie Chaplin, actor, producer, director, and studio founder

1921

"The wireless music box has no imaginable commercial value. Who would pay for a message sent to no one in particular?"

Associates of commercial radio and television pioneer, David Sarnoff, responding to his call for investment in the radio

1946

"Television won't last because people will soon get tired of staring at a plywood box every night."

Darryl Zanuck, movie producer, 20th Century Fox

1977

"There is no reason for any individual to have a computer in his home."

Ken Olson, president, chairman and founder of Digital Equipment Corporation

1995

"The truth is no online database will replace your daily newspaper…"

Clifford Stoll, Newsweek article entitled ***The Internet? Bah!***

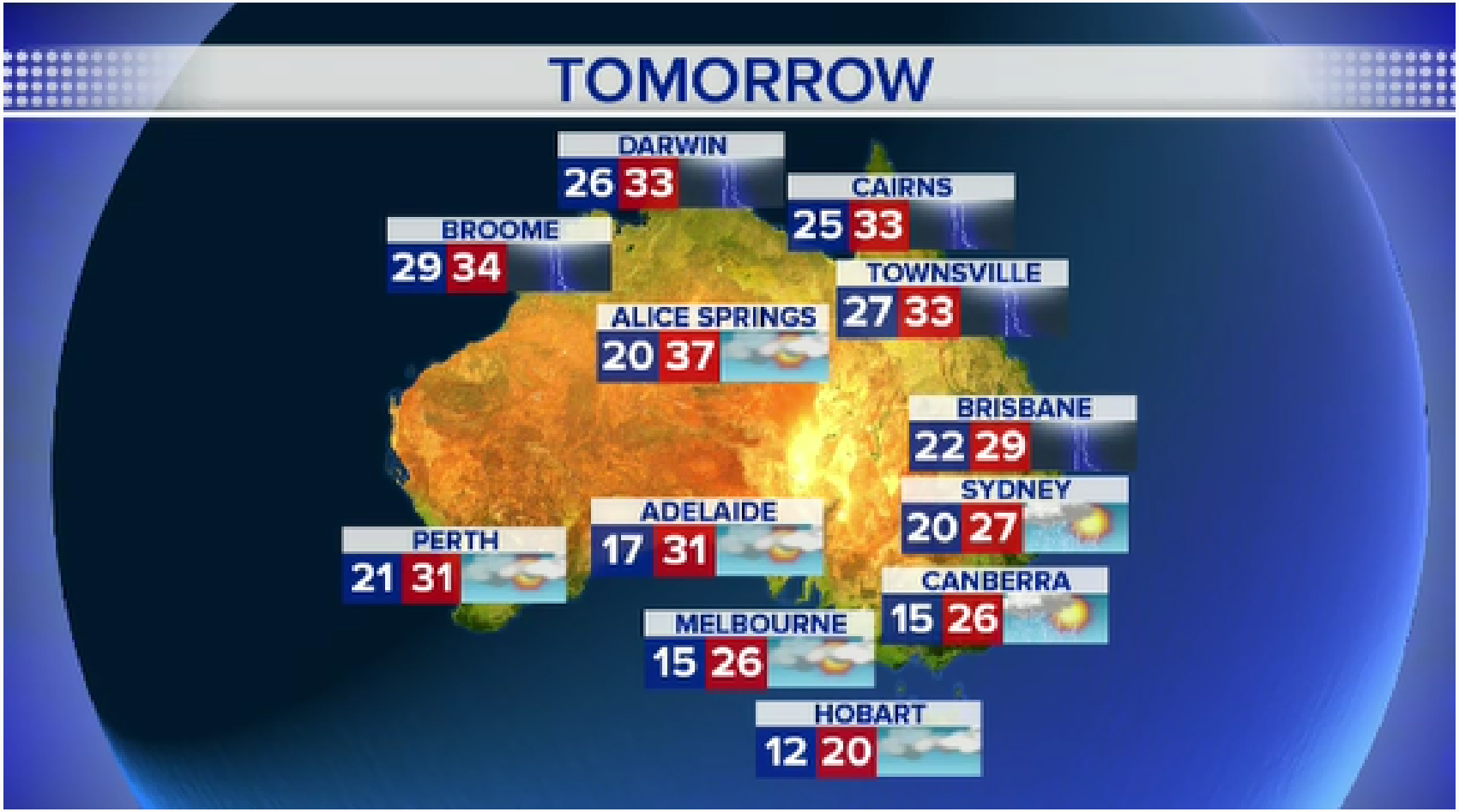
What can we forecast?

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?

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

Which is easiest to forecast?

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

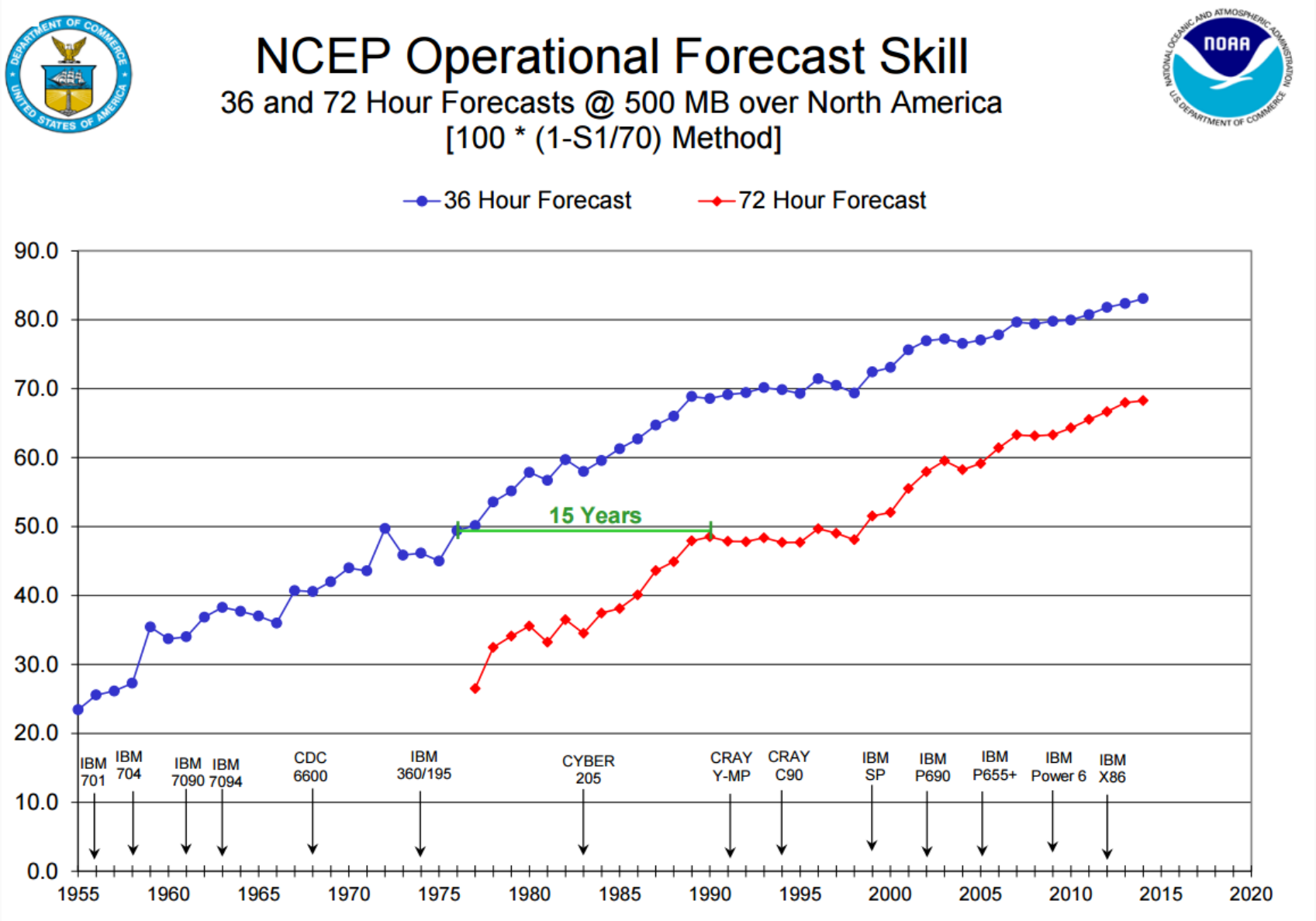
- how do we measure "easiest"?
- what makes something easy/difficult to forecast?

Factors affecting forecastability

Something is easier to forecast if:

- we have a good understanding of the factors that contribute to it
- there is lots of data available;
- the forecasts cannot affect the thing we are trying to forecast.
- there is relatively low natural/unexplainable random variation.
- the future is somewhat similar to the past

Improving forecasts



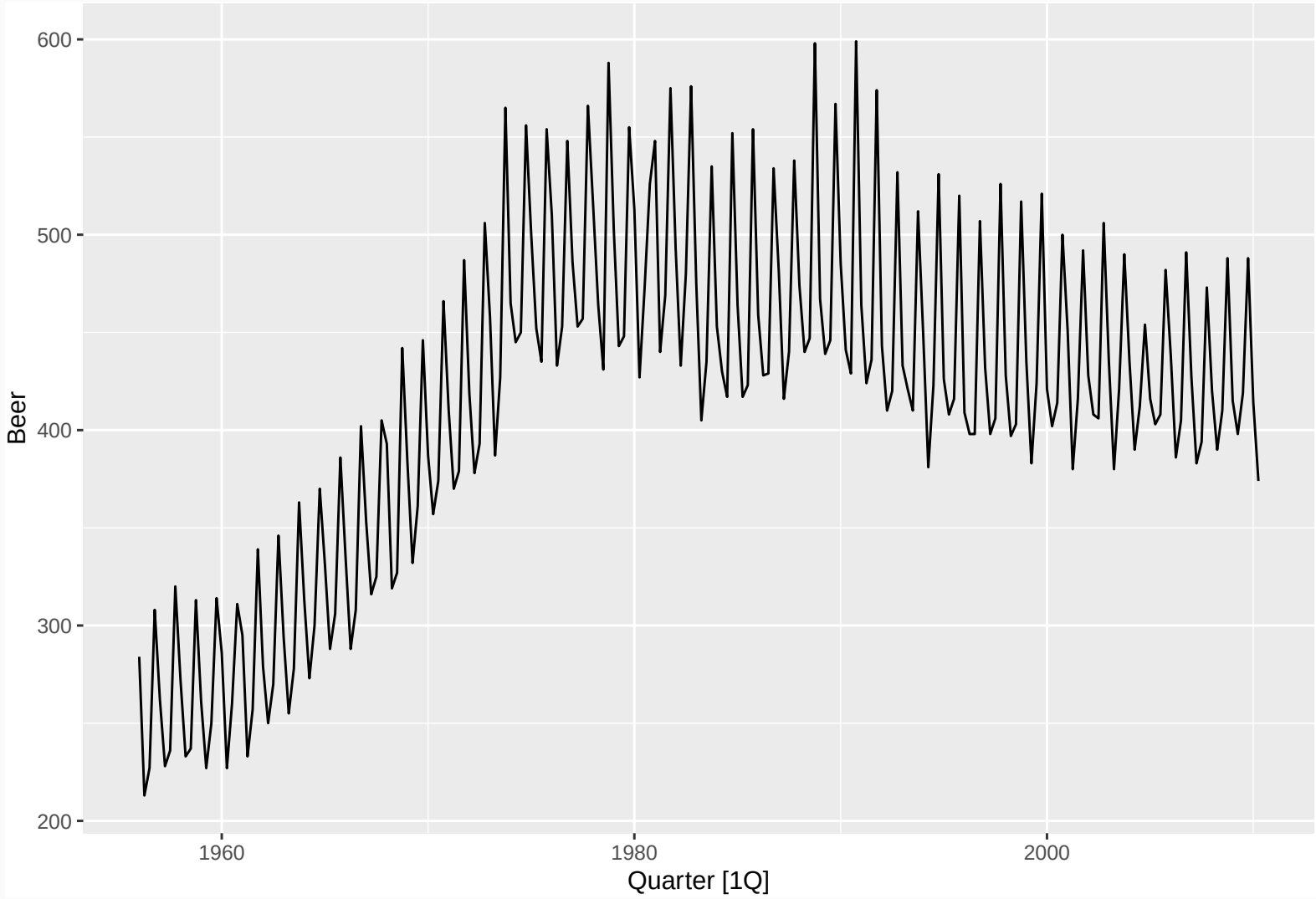
Outline

1. What can we forecast?
2. Time series data
3. Some case studies
4. The statistical forecasting perspective
5. Assignment 1

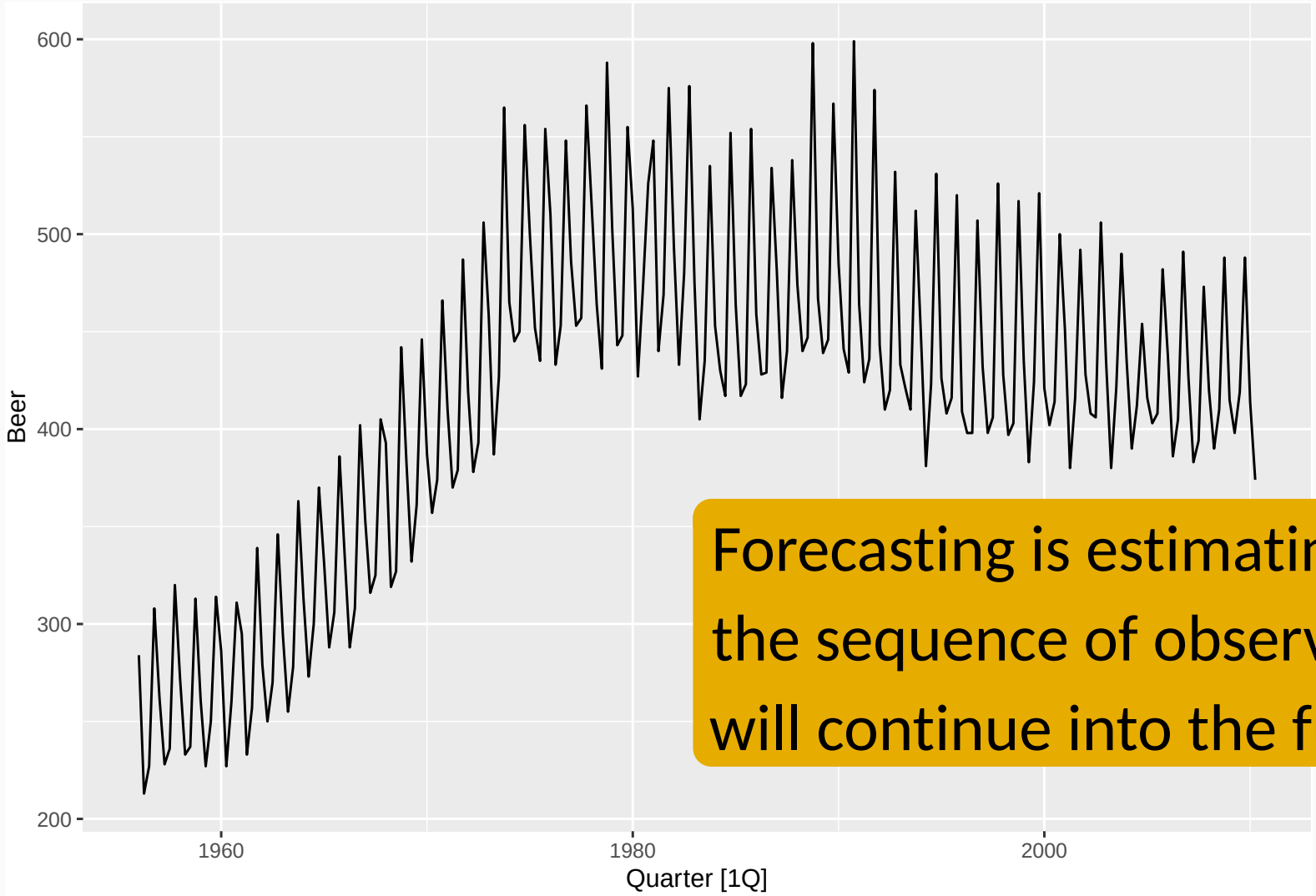
Time series data

- 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

Australian quarterly beer production

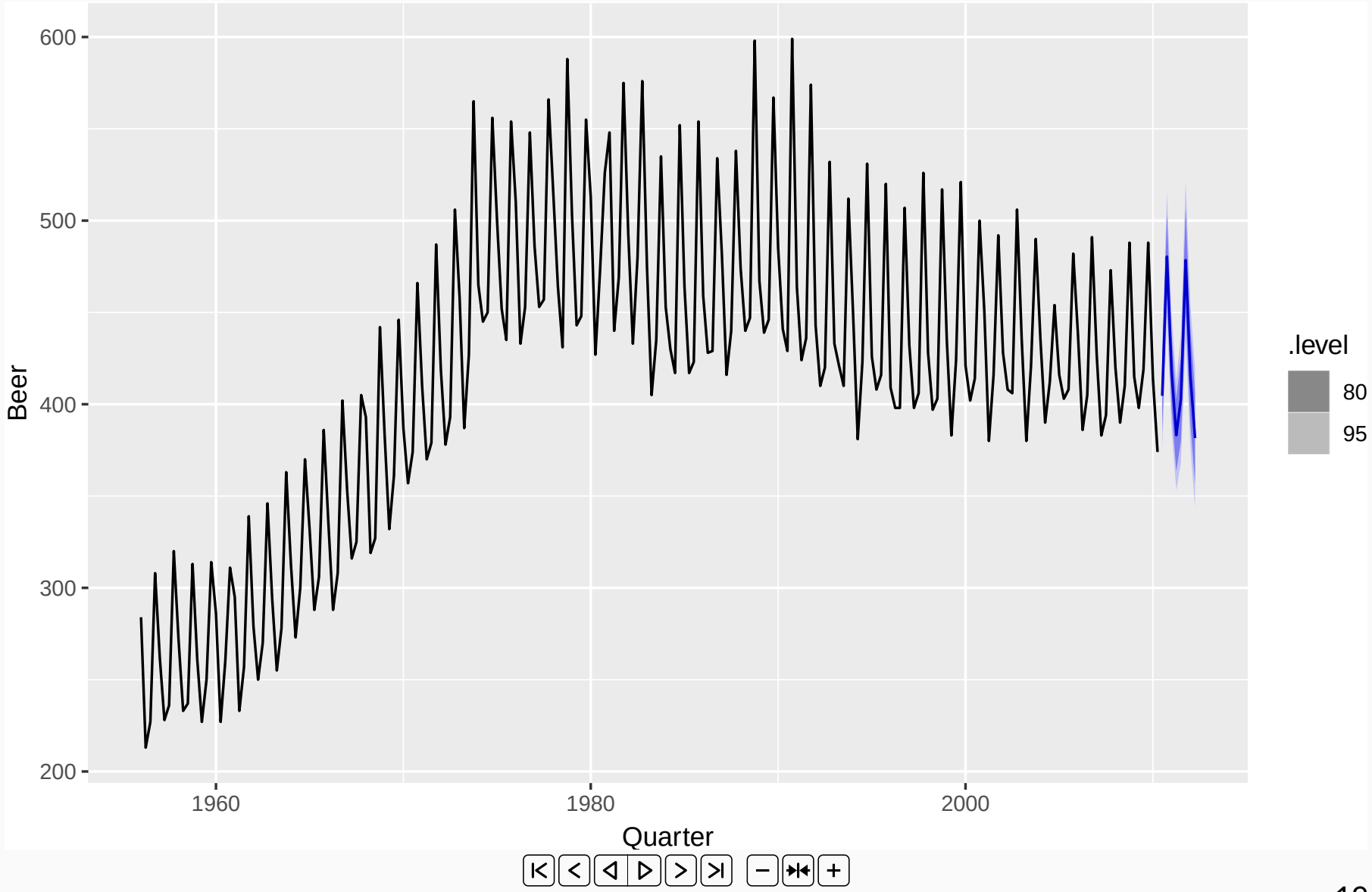


Australian quarterly beer production



Forecasting is estimating how the sequence of observations will continue into the future.

Australian quarterly beer production



Outline

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 tell them what is wrong and 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

CASE STUDY 1: Paperware company

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

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.

CASE STUDY 2: PBS

CASE STUDY 2: PBS

- 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

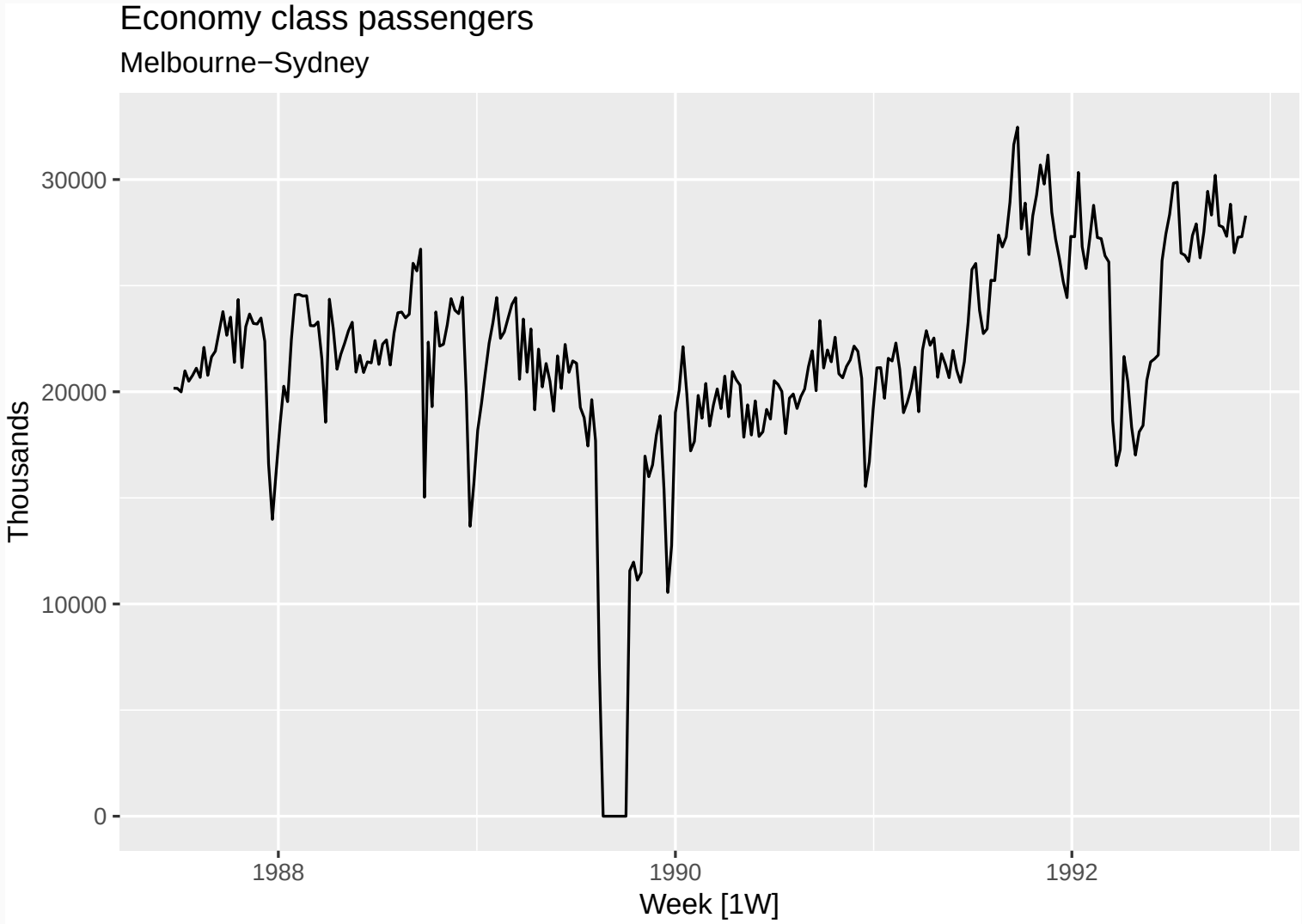
Problem: how to forecast resale value of vehicles? How should this affect leasing and sales policies?

Additional information

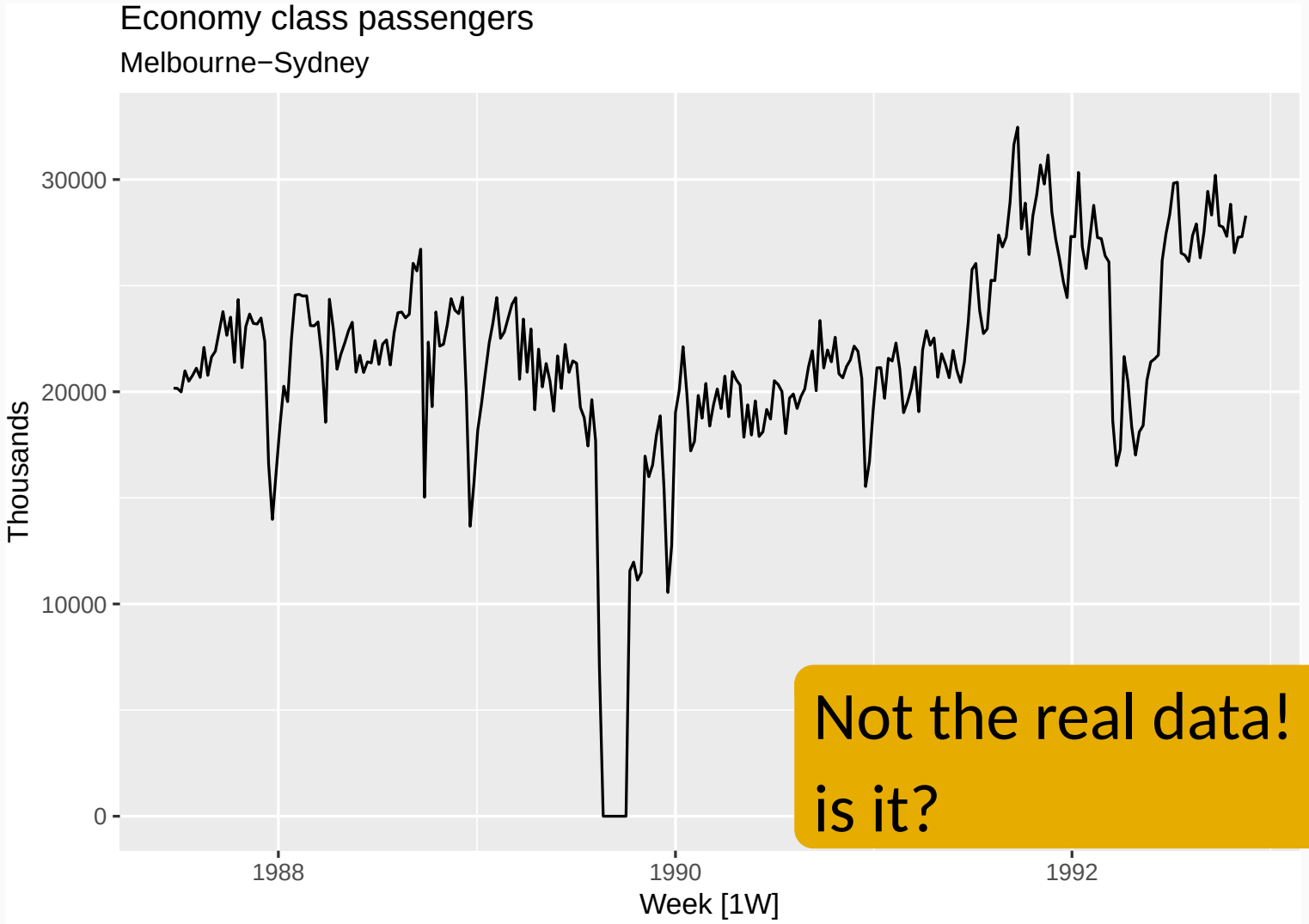
- They can provide a large amount of data on previous vehicles and their eventual resale values.
- The resale values are currently estimated by a group of specialists. They see me as a threat and do not cooperate.

CASE STUDY 4: Airline

CASE STUDY 4: Airline



CASE STUDY 4: Airline



Not the real data! Or is it?

CASE STUDY 4: Airline

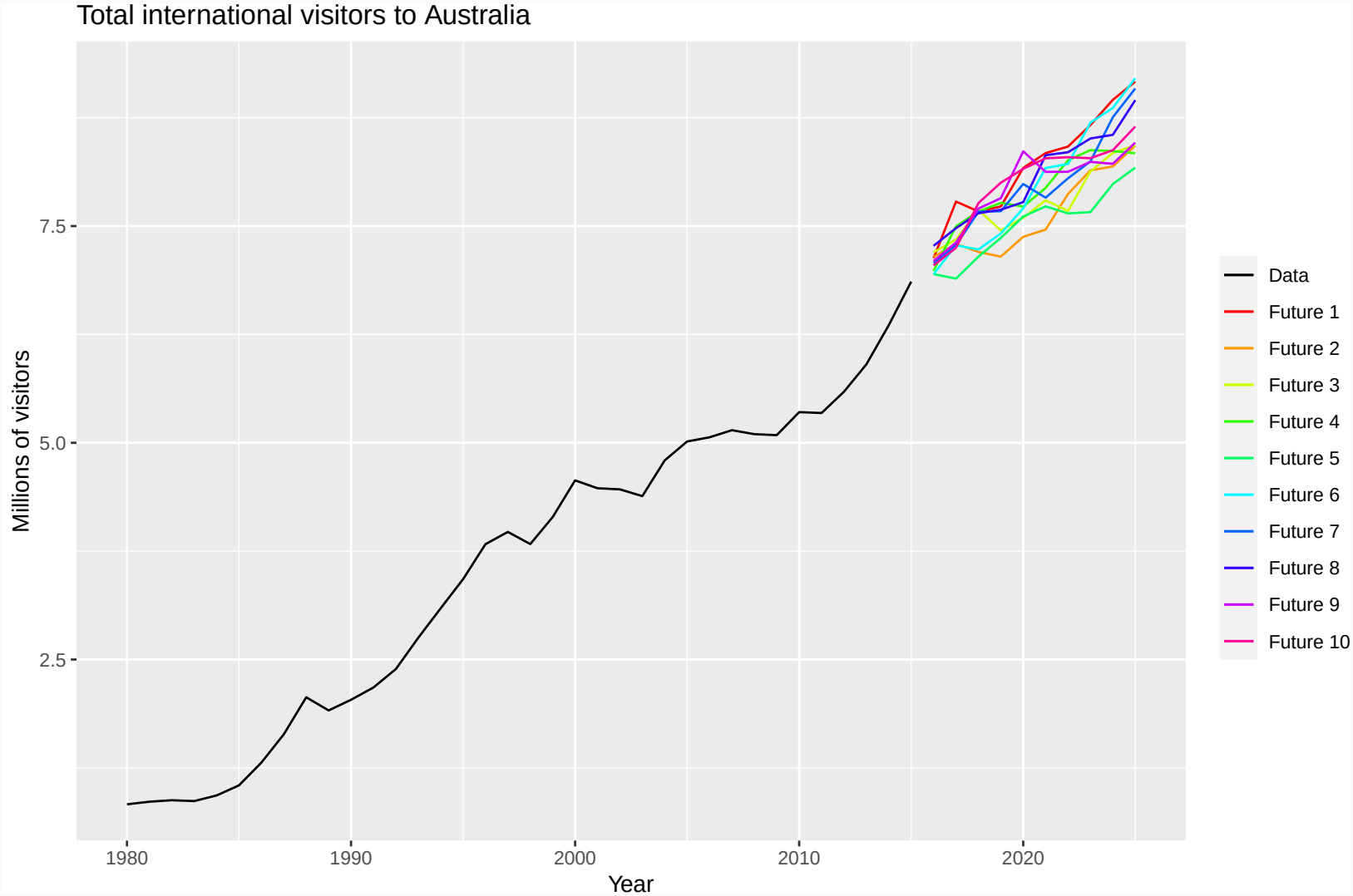
Problem: how to forecast passenger traffic on major routes?

Additional information

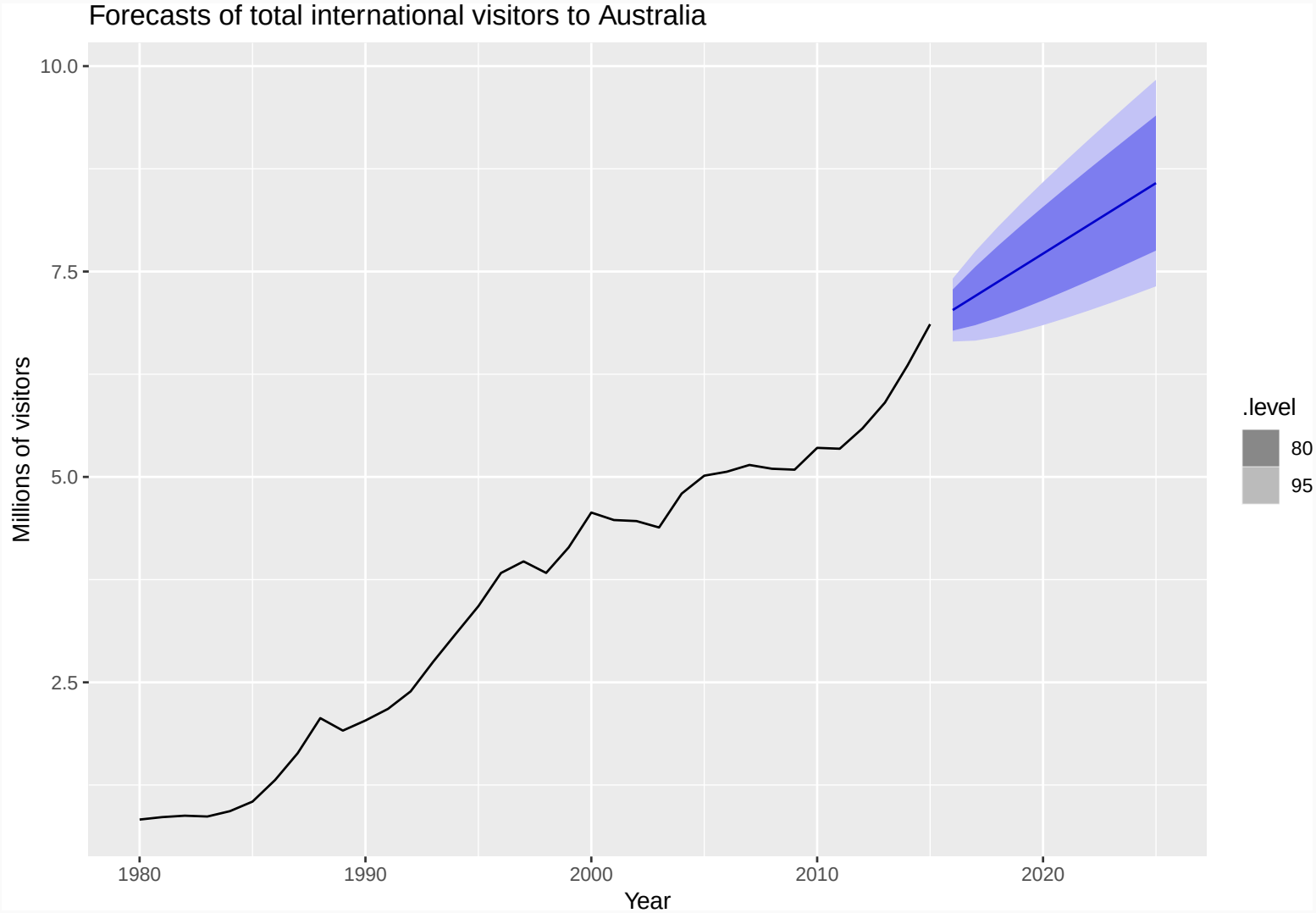
- They can provide a large amount of data on previous routes.
- Traffic is affected by school holidays, special events such as the Grand Prix, advertising campaigns, competition behaviour, etc.
- They have a highly capable team of people who are able to do most of the computing.

Outline

Sample futures



Forecast intervals



Statistical forecasting

- 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, $y_{t|t-1} = y_t|\{y_1, y_2, \ldots, y_{t-1}\}$.
- $\hat{y}_{T+h|T} = \text{E}[y_{T+h}|y_1, \ldots, y_T]$ (an h-step forecast taking account of all observations up to time T).

Outline

1. What can we forecast?
2. Time series data
3. Some case studies
4. The statistical forecasting perspective
5. Assignment 1

Assignment 1: forecast the following series

1. Google closing stock price on 18 March 2020.
2. Maximum temperature at Melbourne airport on 7 April 2020.
3. The difference in points (Collingwood minus Essendon) scored in the AFL match between Collingwood and Essendon for the Anzac Day clash. 25 April 2020.
4. The trend estimate of total employment for April 2020. ABS CAT 6202, to be released around mid May 2020.
5. Google closing stock price on 27 May 2020.

For each of these, give a point forecast and an 80% prediction interval.

Assignment 1: forecast the following series

1. Google closing stock price on 18 March 2020.
2. Maximum temperature at Melbourne airport on 7 April 2020.
3. The difference in points (Collingwood minus Essendon) scored in the AFL match between Collingwood and Essendon for the Anzac Day clash. 25 April 2020.
4. The trend estimate of total employment for April 2020. ABS CAT 6202, to be released around mid May 2020.
5. Google closing stock price on 27 May 2020.

For each of these, give a point forecast and an 80% prediction interval.

Prize: $50 Amazon gift voucher

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