MONASH University

MONASH BUSINESS SCHOOL

# Forecasting using R

**Rob J Hyndman**

Introduction to Forecasting

# Resources

- Slides
- Exercises
- Textbook
- Useful links

**robjhyndman.com/eindhoven**

# Brief bio

- Professor of Statistics, Monash University
- Editor-in-Chief, *International Journal of Forecasting*

**How my forecasting methodology is used:**

- Pharmaceutical Benefits Scheme
- Cancer incidence and mortality
- Electricity demand
- Ageing population
- Fertilizer sales

robjhyndman.com

# Brief bio

- Professor of Statistics, Monash University
- Editor-in-Chief, *International Journal of Forecasting*

## How my forecasting methodology is used:

- Pharmaceutical Benefits Scheme
- Cancer incidence and mortality
- Electricity demand
- Ageing population
- Fertilizer sales

robjhyndman.com

# Assumptions

- This is not an introduction to R. I assume you are broadly comfortable with R code and the RStudio environment.
- This is not a statistics course. I assume you are familiar with concepts such as the mean, standard deviation, quantiles, regression, normal distribution, likelihood, etc.
- This is not a theory course. I am not going to derive anything. I will teach you forecasting tools, when to use them and how to use them most effectively.

# Key reference

**Hyndman, R. J. & Athanasopoulos, G. (2014)**
***Forecasting: principles and practice*.**

**OTexts.org/fpp/**

- Free and online
- Data sets in associated R package
- R code for examples

# Key reference

**Hyndman, R. J. & Athanasopoulos, G. (2014)**
***Forecasting: principles and practice*.**

**OTexts.org/fpp/**

- Free and online
- Data sets in associated R package
- R code for examples

# Key reference

**Hyndman, R. J. & Athanasopoulos, G. (2014)**
***Forecasting: principles and practice*.**

**OTexts.org/fpp/**

- Free and online
- Data sets in associated R package
- R code for examples

# forecasting

1. Guru: I wrote the book, done it for decades, now I do the conference circuit.
2. Expert: It has been my full time job for more than a decade.
3. Skilled: I have been doing it for years.
4. Comfortable: I understand it and have done it.
5. Learner: I am still learning.
6. Beginner: I have heard of it and would like to learn more.
7. Unknown: What is forecasting? Is that what the weather people do?

# Poll: How proficient are you in using R?

1. Guru: The R core team come to me for advice.
2. Expert: I have written several packages on CRAN.
3. Skilled: I use it regularly and it is an important part of my job.
4. Comfortable: I use it often and am comfortable with the tool.
5. User: I use it sometimes, but I am often searching around for the right function.
6. Learner: I have used it a few times.
7. Beginner: I've managed to download and install it.
8. Unknown: Why are you speaking like a pirate?

# Which version of R are you using?

## Version:

(try `getRversion()` if you don't know)

- R 3.3.1
- R 3.2.x
- R 2.15.x
- Something older.

## Edition

- Standard R
- Standard R with RStudio
- Revolution R: Community, Enterprise Workstation or Server
- Something else?

# Install required packages

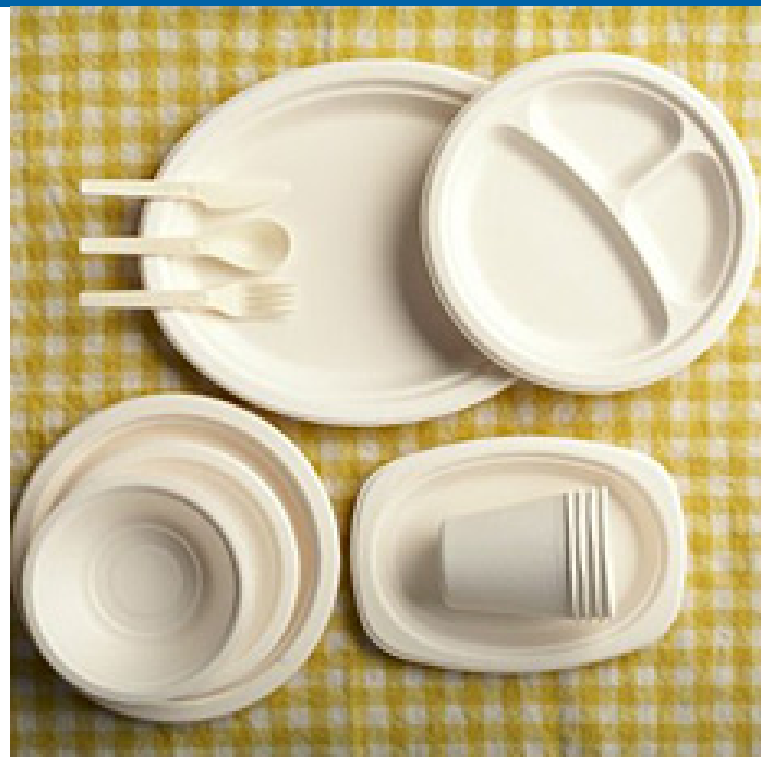
```
install.packages("fpp", dependencies=TRUE)
```

# Approximate outline

| Day | Topic | Chapter |
|---|---|---|
| 1.1 | Time series graphics | 2 |
| 1.2 | The forecaster's toolbox | 2 |
| 1.3 | Seasonality and trends | 6 |
| 1.4 | Exponential smoothing | 7 |
| 2.1 | ETS state space models | 7 |
| 2.2 | Transformations | 2 |
| 2.3 | Stationarity and differencing | 8 |
| 2.4 | Non-seasonal ARIMA models | 8 |
| 2.5 | Seasonal ARIMA models | 8 |
| 3.1 | Time series cross-validation | 2 |
| 3.2 | Dynamic regression | 9 |
| 3.3 | Hierarchical forecasting | 9 |
| 3.4 | Advanced models | 9 |

# 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 the program to produce forecasts automatically.

# 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.

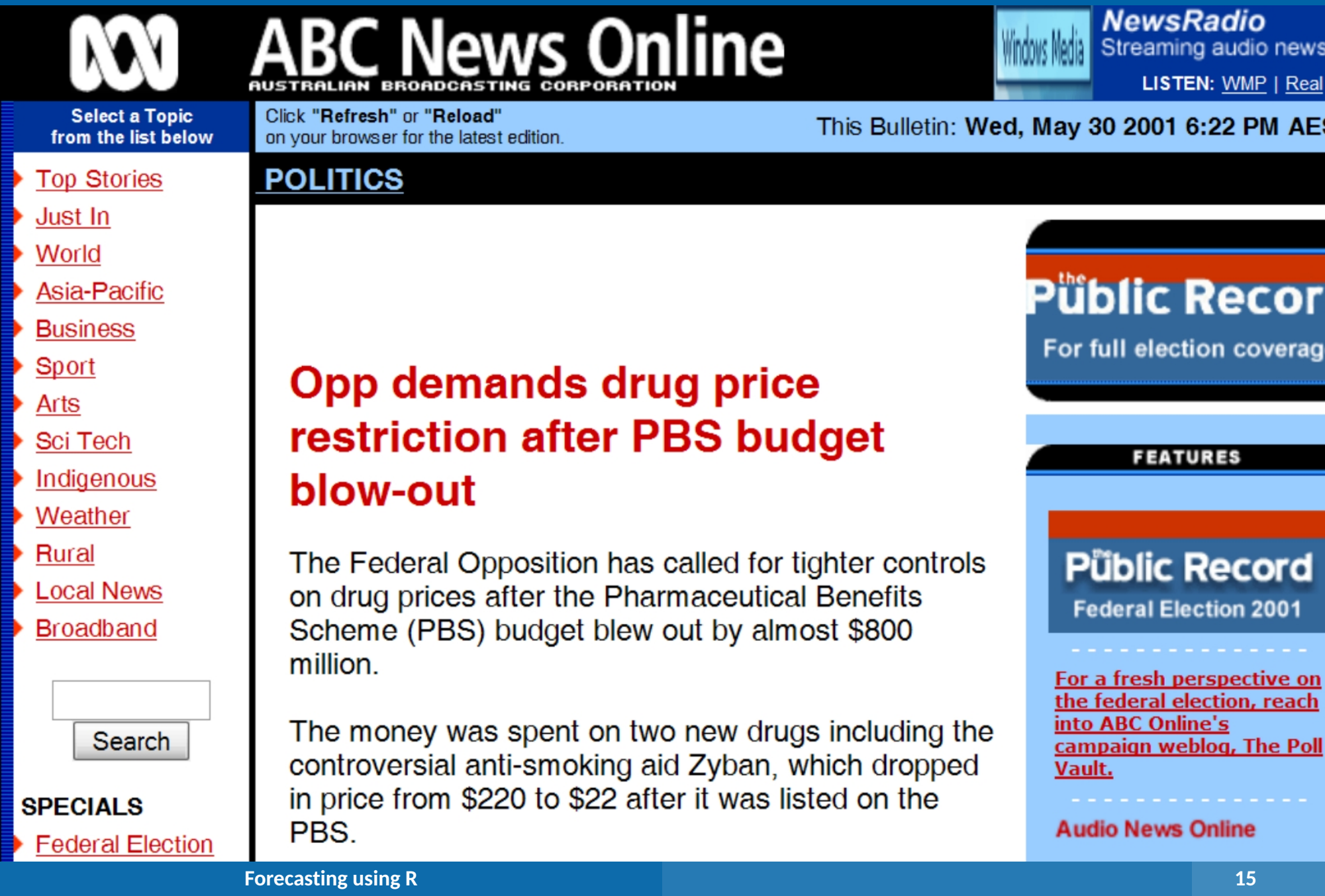
- Many drugs bought from pharmacies are subsidised to allow more equitable access to modern drugs.
- The cost to government is determined by the number and types of drugs purchased. Currently nearly 1% of GDP.
- The total cost is budgeted based on forecasts of drug usage.

# CASE STUDY 2: PBS

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!

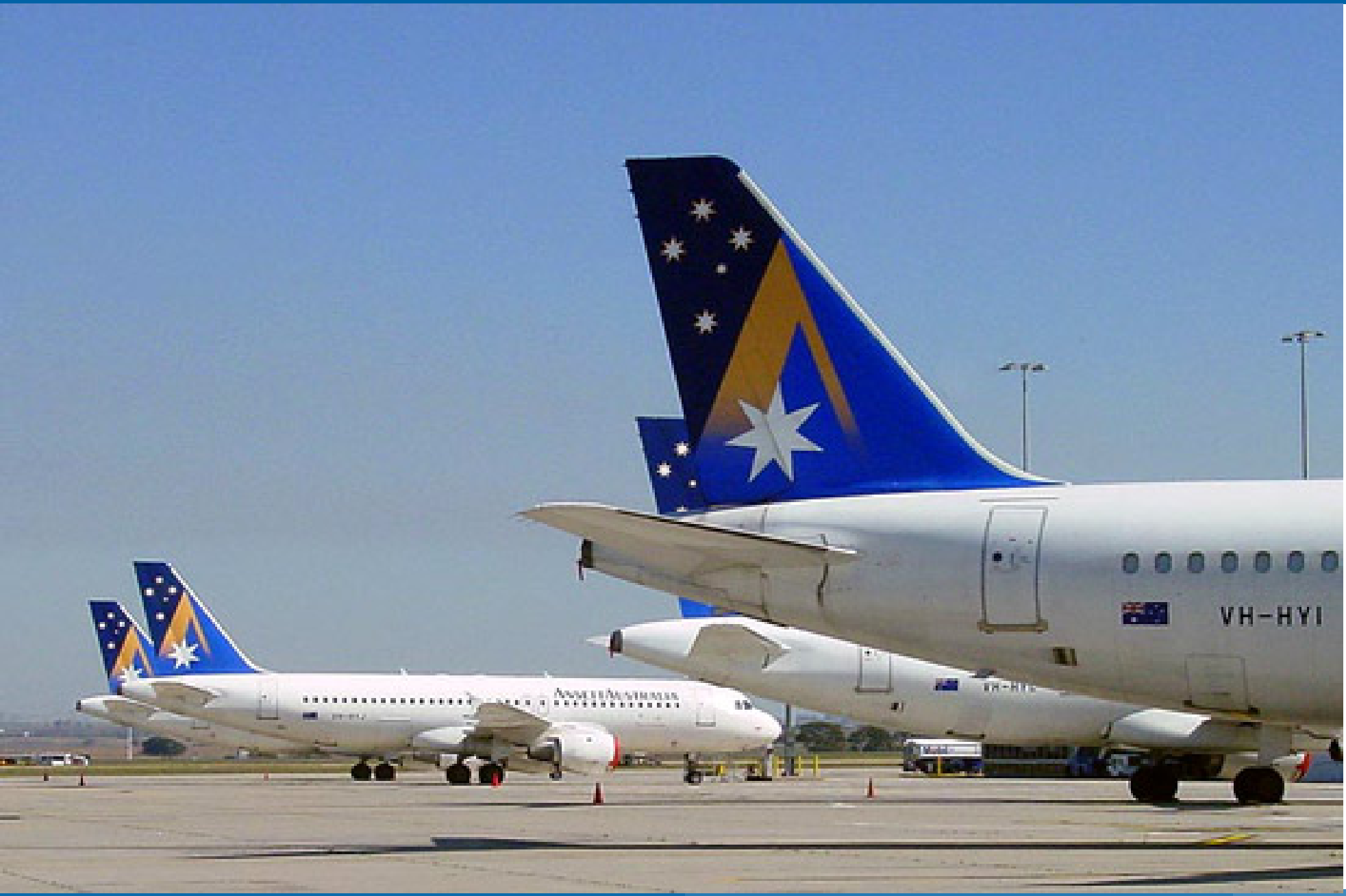
**Problem:** How to do the forecasting better?

# CASE STUDY 2: PBS

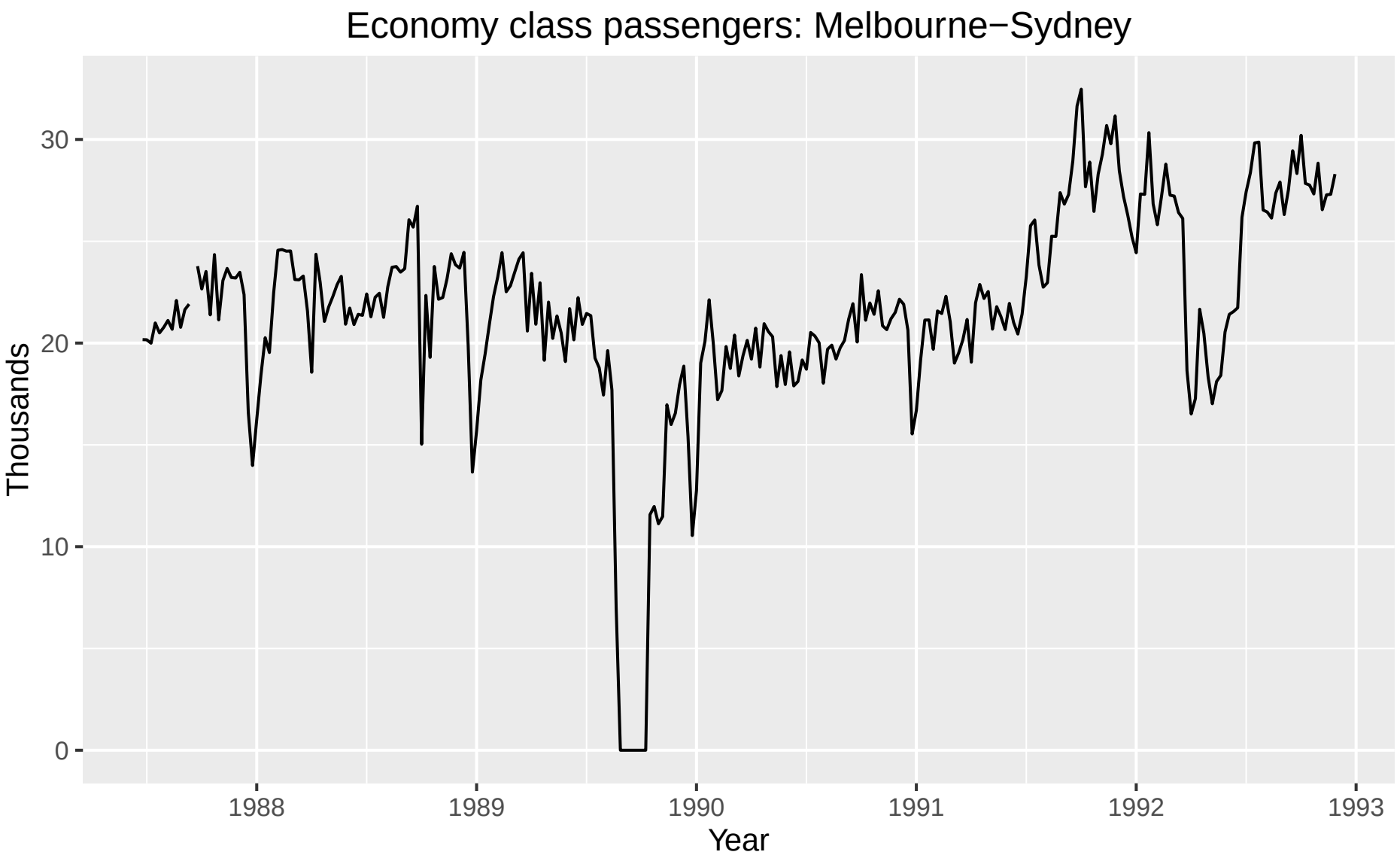
- In 2001: $4.5 billion budget, under-forecasted by $800~million.
- Thousands of products. Seasonal demand.
- Subject to covert marketing, volatile products, uncontrollable expenditure.
- Although monthly data available for 10 years, data are aggregated to annual values, and only the first three years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

**Problem:** How to do the forecasting better?

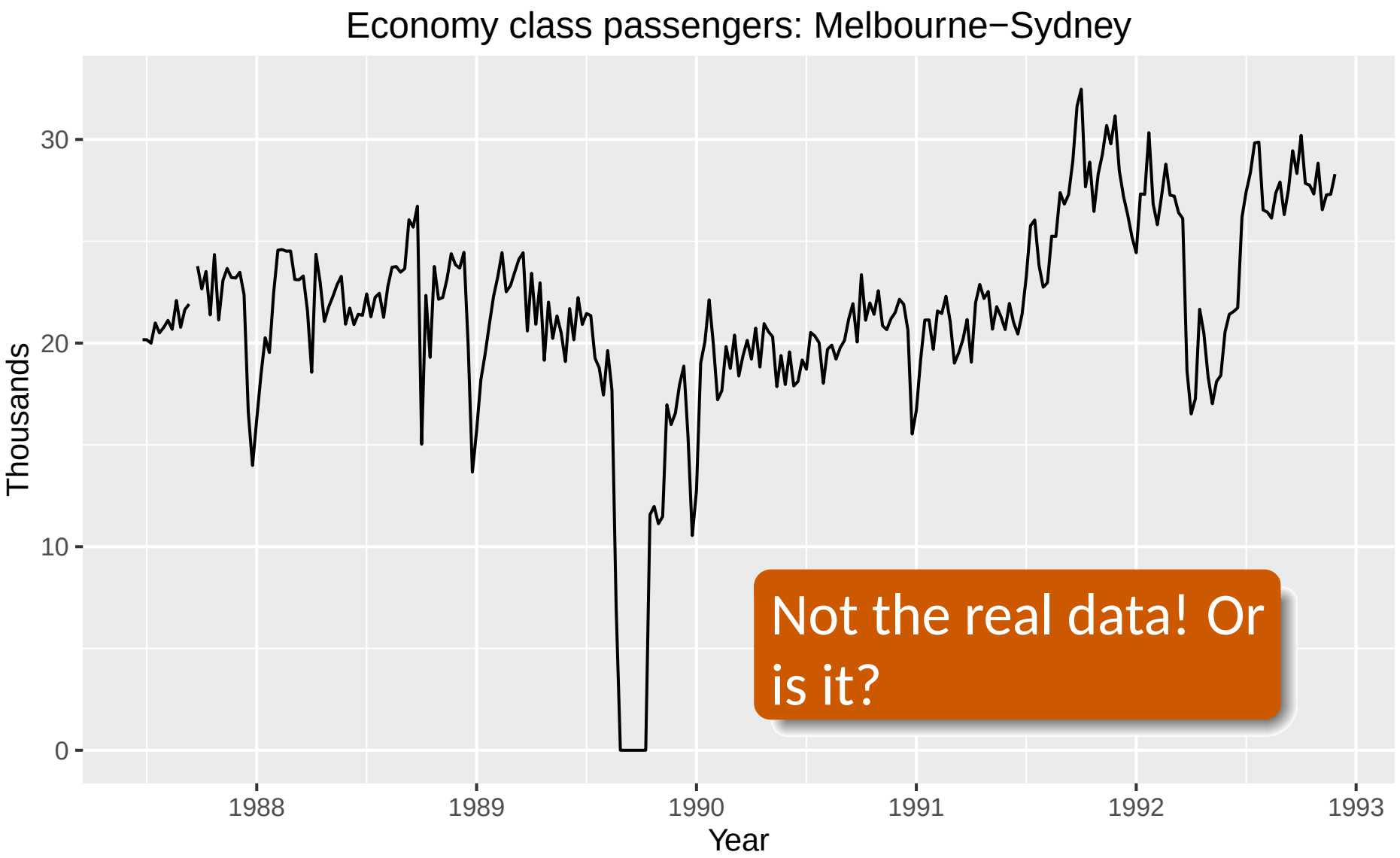
# CASE STUDY 3: Airline



# CASE STUDY 3: Airline



# CASE STUDY 3: Airline



# CASE STUDY 3: 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.