

ETC3550/ETC5550

Applied forecasting

Week 1: Introduction to forecasting & R



Contact details

Chief Examiner: Professor Rob Hyndman

✉ rob.hyndman@monash.edu

🏠 robjhyndman.com

Tutors

- **Mitchell O'Hara-Wild**
- Maliny Po
- Nuwani Palihawadana
- Xiefei (Sapphire) Li

Brief bio

- Professor of Statistics, Monash University
- Co-author of most popular forecasting textbook in the world
- Lead developer of most popular forecasting software in the world

How my forecasting methodology is used:

- Pharmaceutical Benefits Scheme
- Electricity demand
- Australian tourism demand
- Ageing population
- COVID-19 cases

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.



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



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

the Public Record

For full election coverage

FEATURES

the Public Record

Federal Election 2001

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

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



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.

Unit objectives

- 1 To obtain an understanding of common statistical methods used in business and economic forecasting.
- 2 To develop the computer skills required to forecast business and economic time series data;
- 3 To gain insights into the problems of implementing and operating large scale forecasting systems for use in business.

Unit objectives

- 1 To obtain an understanding of common statistical methods used in business and economic forecasting.
- 2 To develop the computer skills required to forecast business and economic time series data;
- 3 To gain insights into the problems of implementing and operating large scale forecasting systems for use in business.

Teaching and learning approach

- Approximately one hour of pre-recorded online videos each week
- One hour online lecture each Monday
- One hour in-person workshop each Tuesday (focus on exam)
- One hour in-person tutorial each week (focus on assignments)

Key reference

Hyndman, R. J. & Athanasopoulos, G. (2021) *Forecasting: principles and practice*, 3rd edition

Key reference

Hyndman, R. J. & Athanasopoulos, G. (2021) *Forecasting: principles and practice*, 3rd edition

[OTexts.com/fpp3/](https://otexts.com/fpp3/)

Key reference

Hyndman, R. J. & Athanasopoulos, G. (2021) *Forecasting: principles and practice*, 3rd edition

[OTexts.com/fpp3/](https://otexts.com/fpp3/)

- Free and online
- Data sets in associated R packages
- R code for examples
- Embedded online lectures

Outline

| Week | Topic | Chapter |
|------|-------------------------------------|---------|
| 1 | Introduction to forecasting and R | 1 |
| 2 | Time series graphics | 2 |
| 3 | Time series decomposition | 3 |
| 4 | Simple forecasting methods | 5 |
| 5 | Accuracy evaluation | 5 |
| 6-7 | Exponential smoothing | 8 |
| 8-10 | ARIMA models | 9 |
| 11 | Multiple regression and forecasting | 7 |
| 12 | Dynamic regression | 10 |

Assessment

| Task | Due Date | Value |
|-------------------------|----------------------|-------|
| Forecasting Competition | Fri 7 Mar | 2% |
| Weekly Quizzes | end of weeks 2–11 | 8% |
| Assignment 1 | Fri 28 Mar | 6% |
| Assignment 2 | Fri 25 Apr | 6% |
| Assignment 3 | Tue 20 May | 6% |
| Retail Project | Fri 30 May | 12% |
| Final Exam | Official exam period | 60% |

Assessment

| Task | Due Date | Value |
|-------------------------|----------------------|-------|
| Forecasting Competition | Fri 7 Mar | 2% |
| Weekly Quizzes | end of weeks 2–11 | 8% |
| Assignment 1 | Fri 28 Mar | 6% |
| Assignment 2 | Fri 25 Apr | 6% |
| Assignment 3 | Tue 20 May | 6% |
| Retail Project | Fri 30 May | 12% |
| Final Exam | Official exam period | 60% |

- Need at least 45% for exam, and 50% for total.
- **ETC5550 students:** Extra project and exam questions.

af.numbat.space

- Includes all course materials
- Links for assignment submissions
- Link to discussion forum.

Please don't send emails. Use the forum.



- The IIF provides a prize to the top student in this subject each year.
- US\$100 plus one year membership.



Available for download from CRAN:

<https://cran.r-project.org>



Available for download from RStudio:

<https://www.rstudio.com/products/rstudio/download/>

Main packages



Main packages

```
# Install required packages (do once)
install.packages(c("tidyverse", "fpp3", "GGally"), dependencies = TRUE)
```

Main packages

```
# Install required packages (do once)  
install.packages(c("tidyverse", "fpp3", "GGally), dependencies = TRUE)
```

```
# At the start of each session  
library(fpp3)
```

Exercises Week 1

- Make sure you are familiar with R, RStudio and the tidyverse packages.
- Do the first five modules of `startr.numbat.space`.
- Forecasting competition

Forecasting competition: forecast the following series

- 1 Google closing stock price on 24 March 2025
- 2 Maximum temperature at Melbourne airport on 14 April 2025
- 3 The difference in points (Collingwood minus Essendon) scored in the AFL match between Collingwood and Essendon for the Anzac Day clash. 25 April 2025
- 4 The seasonally adjusted estimate of total employment for April 2025. ABS CAT 6202, to be released around mid May 2025
- 5 Google closing stock price on 26 May 2025

Due Friday 14 March

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

Forecasting competition: forecast the following series

- 1 Google closing stock price on 24 March 2025
- 2 Maximum temperature at Melbourne airport on 14 April 2025
- 3 The difference in points (Collingwood minus Essendon) scored in the AFL match between Collingwood and Essendon for the Anzac Day clash. 25 April 2025
- 4 The seasonally adjusted estimate of total employment for April 2025. ABS CAT 6202, to be released around mid May 2025
- 5 Google closing stock price on 26 May 2025

Due Friday 14 March

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

Prize: \$50 cash prize

Forecasting competition: 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:

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

- $u_+ = \max(u, 0)$
- Rank results for all students
- Add ranks across all five items

tsibble objects

```
global_economy
```

```
# A tsibble: 15,150 x 6 [1Y]
```

```
# Key:      Country [263]
```

| | Year | Country | GDP | Imports | Exports | Population |
|----|-------|-------------|--------------|---------|---------|------------|
| | <dbl> | <fct> | <dbl> | <dbl> | <dbl> | <dbl> |
| 1 | 1960 | Afghanistan | 5377777811. | 7.02 | 4.13 | 8996351 |
| 2 | 1961 | Afghanistan | 5488888896. | 8.10 | 4.45 | 9166764 |
| 3 | 1962 | Afghanistan | 5466666678. | 9.35 | 4.88 | 9345868 |
| 4 | 1963 | Afghanistan | 7511111191. | 16.9 | 9.17 | 9533954 |
| 5 | 1964 | Afghanistan | 8000000044. | 18.1 | 8.89 | 9731361 |
| 6 | 1965 | Afghanistan | 10066666638. | 21.4 | 11.3 | 9938414 |
| 7 | 1966 | Afghanistan | 13999999967. | 18.6 | 8.57 | 10152331 |
| 8 | 1967 | Afghanistan | 16733333418. | 14.2 | 6.77 | 10372630 |
| 9 | 1968 | Afghanistan | 13733333367. | 15.2 | 8.90 | 10604346 |
| 10 | 1969 | Afghanistan | 14088888922. | 15.0 | 10.1 | 10854428 |

```
# i 15,140 more rows
```

tsibble objects

```
global_economy
```

```
# A tsibble: 15,150 x 6 [1Y]
```

```
# Key:           Country [263]
```

| | Year | Country | GDP | Imports | Exports | Population |
|----|-------|-------------|--------------|---------|---------|------------|
| | Index | <fct> | <dbl> | <dbl> | <dbl> | <dbl> |
| 1 | 1960 | Afghanistan | 5377777811. | 7.02 | 4.13 | 8996351 |
| 2 | 1961 | Afghanistan | 5488888896. | 8.10 | 4.45 | 9166764 |
| 3 | 1962 | Afghanistan | 5466666678. | 9.35 | 4.88 | 9345868 |
| 4 | 1963 | Afghanistan | 7511111191. | 16.9 | 9.17 | 9533954 |
| 5 | 1964 | Afghanistan | 8000000044. | 18.1 | 8.89 | 9731361 |
| 6 | 1965 | Afghanistan | 10066666638. | 21.4 | 11.3 | 9938414 |
| 7 | 1966 | Afghanistan | 13999999967. | 18.6 | 8.57 | 10152331 |
| 8 | 1967 | Afghanistan | 16733333418. | 14.2 | 6.77 | 10372630 |
| 9 | 1968 | Afghanistan | 13733333367. | 15.2 | 8.90 | 10604346 |
| 10 | 1969 | Afghanistan | 14088888922. | 15.0 | 10.1 | 10854428 |

```
# i 15,140 more rows
```

tsibble objects

```
global_economy
```

```
# A tsibble: 15,150 x 6 [1Y]
```

```
# Key:      Country [263]
```

| | Year | Country | GDP | Imports | Exports | Population |
|----|-------|-------------|--------------|---------|---------|------------|
| | Index | Key | <dbl> | <dbl> | <dbl> | <dbl> |
| 1 | 1960 | Afghanistan | 5377777811. | 7.02 | 4.13 | 8996351 |
| 2 | 1961 | Afghanistan | 5488888896. | 8.10 | 4.45 | 9166764 |
| 3 | 1962 | Afghanistan | 5466666678. | 9.35 | 4.88 | 9345868 |
| 4 | 1963 | Afghanistan | 7511111191. | 16.9 | 9.17 | 9533954 |
| 5 | 1964 | Afghanistan | 8000000044. | 18.1 | 8.89 | 9731361 |
| 6 | 1965 | Afghanistan | 10066666638. | 21.4 | 11.3 | 9938414 |
| 7 | 1966 | Afghanistan | 13999999967. | 18.6 | 8.57 | 10152331 |
| 8 | 1967 | Afghanistan | 16733333418. | 14.2 | 6.77 | 10372630 |
| 9 | 1968 | Afghanistan | 13733333367. | 15.2 | 8.90 | 10604346 |
| 10 | 1969 | Afghanistan | 14088888922. | 15.0 | 10.1 | 10854428 |

```
# i 15,140 more rows
```

tsibble objects

```
global_economy
```

```
# A tsibble: 15,150 x 6 [1Y]
```

```
# Key:      Country [263]
```

| | Year | Country | GDP | Imports | Exports | Population |
|----|-------|-------------|--------------------|---------|---------|------------|
| | Index | Key | Measured variables | | | |
| 1 | 1960 | Afghanistan | 537777811. | 7.02 | 4.13 | 8996351 |
| 2 | 1961 | Afghanistan | 548888896. | 8.10 | 4.45 | 9166764 |
| 3 | 1962 | Afghanistan | 546666678. | 9.35 | 4.88 | 9345868 |
| 4 | 1963 | Afghanistan | 751111191. | 16.9 | 9.17 | 9533954 |
| 5 | 1964 | Afghanistan | 800000044. | 18.1 | 8.89 | 9731361 |
| 6 | 1965 | Afghanistan | 1006666638. | 21.4 | 11.3 | 9938414 |
| 7 | 1966 | Afghanistan | 1399999967. | 18.6 | 8.57 | 10152331 |
| 8 | 1967 | Afghanistan | 1673333418. | 14.2 | 6.77 | 10372630 |
| 9 | 1968 | Afghanistan | 1373333367. | 15.2 | 8.90 | 10604346 |
| 10 | 1969 | Afghanistan | 1408888922. | 15.0 | 10.1 | 10854428 |

```
# i 15,140 more rows
```

tsibble objects

```
tourism
```

```
# A tsibble: 24,320 x 5 [1Q]
# Key:           Region, State, Purpose [304]
  Quarter Region  State Purpose  Trips
   <qtr>  <chr>    <chr> <chr>    <dbl>
1 1998 Q1 Adelaide SA      Business 135.
2 1998 Q2 Adelaide SA      Business 110.
3 1998 Q3 Adelaide SA      Business 166.
4 1998 Q4 Adelaide SA      Business 127.
5 1999 Q1 Adelaide SA      Business 137.
6 1999 Q2 Adelaide SA      Business 200.
7 1999 Q3 Adelaide SA      Business 169.
8 1999 Q4 Adelaide SA      Business 134.
9 2000 Q1 Adelaide SA      Business 154.
10 2000 Q2 Adelaide SA      Business 169.
# i 24,310 more rows
```

tsibble objects

```
tourism
```

```
# A tsibble: 24,320 x 5 [1Q]
# Key:           Region, State, Purpose [304]
   Quarter Region  State Purpose  Trips
   <qtr>  <chr>    <chr> <chr>    <dbl>
1 1998 Q1 Adelaide SA      Business 135.
2 1998 Q2 Adelaide SA      Business 110.
3 1998 Q3 Adelaide SA      Business 166.
4 1998 Q4 Adelaide SA      Business 127.
5 1999 Q1 Adelaide SA      Business 137.
6 1999 Q2 Adelaide SA      Business 200.
7 1999 Q3 Adelaide SA      Business 169.
8 1999 Q4 Adelaide SA      Business 134.
9 2000 Q1 Adelaide SA      Business 154.
10 2000 Q2 Adelaide SA      Business 169.
# i 24,310 more rows
```

Domestic visitor
nights in
thousands by
state/region and
purpose.

tsibble objects

```
tourism
```

```
# A tsibble: 24,320 x 5 [1Q]
# Key:           Region, State, Purpose [304]
  Quarter Region State Purpose Trips
  <dbl>   <chr>   <chr> <chr>   <dbl>
1 1998 Q1 Adelaide SA      Business 135.
2 1998 Q2 Adelaide SA      Business 110.
3 1998 Q3 Adelaide SA      Business 166.
4 1998 Q4 Adelaide SA      Business 127.
5 1999 Q1 Adelaide SA      Business 137.
6 1999 Q2 Adelaide SA      Business 200.
7 1999 Q3 Adelaide SA      Business 169.
8 1999 Q4 Adelaide SA      Business 134.
9 2000 Q1 Adelaide SA      Business 154.
10 2000 Q2 Adelaide SA      Business 169.
# i 24,310 more rows
```

Domestic visitor
nights in
thousands by
state/region and
purpose.

tsibble objects

```
tourism
```

```
# A tsibble: 24,320 x 5 [1Q]
# Key:           Region, State, Purpose [304]
  Quarter Region State Purpose Trips
  <dbl>   <chr>  <chr>   <chr>   <dbl>
1 1998 Q1 Adelaide SA      Business 135.
2 1998 Q2 Adelaide SA      Business 110.
3 1998 Q3 Adelaide SA      Business 166.
4 1998 Q4 Adelaide SA      Business 127.
5 1999 Q1 Adelaide SA      Business 137.
6 1999 Q2 Adelaide SA      Business 200.
7 1999 Q3 Adelaide SA      Business 169.
8 1999 Q4 Adelaide SA      Business 134.
9 2000 Q1 Adelaide SA      Business 154.
10 2000 Q2 Adelaide SA      Business 169.
# i 24,310 more rows
```

Domestic visitor
nights in
thousands by
state/region and
purpose.

tsibble objects

```
tourism
```

```
# A tsibble: 24,320 x 5 [1Q]  
# Key:      Region, State, Purpose [304]
```

| | Quarter | Region | State | Purpose | Trips |
|----|---------|----------|-------|----------|---------|
| | Index | Keys | | | Measure |
| 1 | 1998 Q1 | Adelaide | SA | Business | 135. |
| 2 | 1998 Q2 | Adelaide | SA | Business | 110. |
| 3 | 1998 Q3 | Adelaide | SA | Business | 166. |
| 4 | 1998 Q4 | Adelaide | SA | Business | 127. |
| 5 | 1999 Q1 | Adelaide | SA | Business | 137. |
| 6 | 1999 Q2 | Adelaide | SA | Business | 200. |
| 7 | 1999 Q3 | Adelaide | SA | Business | 169. |
| 8 | 1999 Q4 | Adelaide | SA | Business | 134. |
| 9 | 2000 Q1 | Adelaide | SA | Business | 154. |
| 10 | 2000 Q2 | Adelaide | SA | Business | 169. |

```
# i 24,310 more rows
```

Domestic visitor
nights in
thousands by
state/region and
purpose.

tsibble objects

- A `tsibble` allows storage and manipulation of multiple time series in R.
- It contains:
 - ▶ An index: time information about the observation
 - ▶ Measured variable(s): numbers of interest
 - ▶ Key variable(s): optional unique identifiers for each series
- It works with tidyverse functions.

The tsibble index

Time index variables can be created with these functions:

| Frequency | Function |
|-----------|----------|
|-----------|----------|

| | |
|-----------|-------------------------------|
| Annual | <code>start:end</code> |
| Quarterly | <code>yearquarter()</code> |
| Monthly | <code>yearmonth()</code> |
| Weekly | <code>yearweek()</code> |
| Daily | <code>as_date(), ymd()</code> |
| Sub-daily | <code>as_datetime()</code> |