Probabilistic forecasts for anomaly detection

Rob J Hyndman

3 July 2024

MONASH University

Outline

1. Australian PBS data
2. Probabilistic forecasts
3. Extreme log scores
4. Using the fable and weird packages
5. Online anomaly detection

Outline

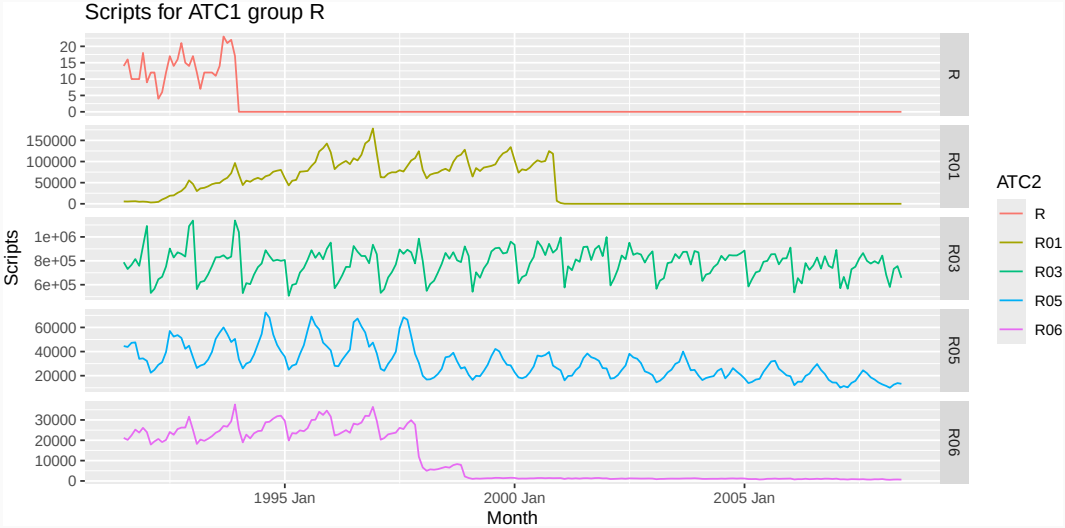
1. Australian PBS data
2. Probabilistic forecasts
3. Extreme log scores
4. Using the fable and weird packages
5. Online anomaly detection

Australian PBS data

```
pbs
```

```
# A tsibble: 17,016 x 4 [1M]
# Key:       ATC1, ATC2 [84]
   ATC1  ATC2     Month Scripts
   <chr> <chr>    <mth>   <dbl>
 1 A     A01   1991 Jul   22615
 2 A     A01   1991 Aug   20443
 3 A     A01   1991 Sep   21389
 4 A     A01   1991 Oct   23746
 5 A     A01   1991 Nov   23477
 6 A     A01   1991 Dec   26316
 7 A     A01   1992 Jan   22041
 8 A     A01   1992 Feb   16393
 9 A     A01   1992 Mar   17207
10 A     A01   1992 Apr   18847
# i 17,006 more rows
```

Australian PBS data



Main idea

- Look at one-step forecast distributions and compute the anomaly score = $-$ log probability density of the next observation.
- High anomaly scores indicate potential anomalies.
- Fit a Generalized Pareto Distribution to the largest anomaly scores.
- Estimate the probability of each observation being an anomaly.

Outline

1. Australian PBS data
2. Probabilistic forecasts
3. Extreme log scores
4. Using the fable and weird packages
5. Online anomaly detection

Probabilistic forecasts

- Describe probabilistic forecasts

Outline

1 Australian PBS data

2 Probabilistic forecasts

3 Extreme log scores

4 Using the fable and weird packages

5 Online anomaly detection

Extreme log scores

- Explain log scores and EVT approach to finding anomalies ala lookout

Outline

1 Australian PBS data

2 Probabilistic forecasts

3 Extreme log scores

4 Using the fable and weird packages

5 Online anomaly detection

Using the fable and weird packages

- Demonstrate using weird package with (a) univariate models for tourism data; and (b) univariate models for age-specific time series from French mortality.

Outline

1 Australian PBS data

2 Probabilistic forecasts

3 Extreme log scores

4 Using the fable and weird packages

5 Online anomaly detection

Online anomaly detection

- Need to compute one-step forecast distributions using prior data. Approximately the same as N(y-hat, sigmaˆ2) using in-sample fit. More accurate using stretch tsibble with one-step forecasts.