Probabilistic forecasts for anomaly detection

Rob J Hyndman

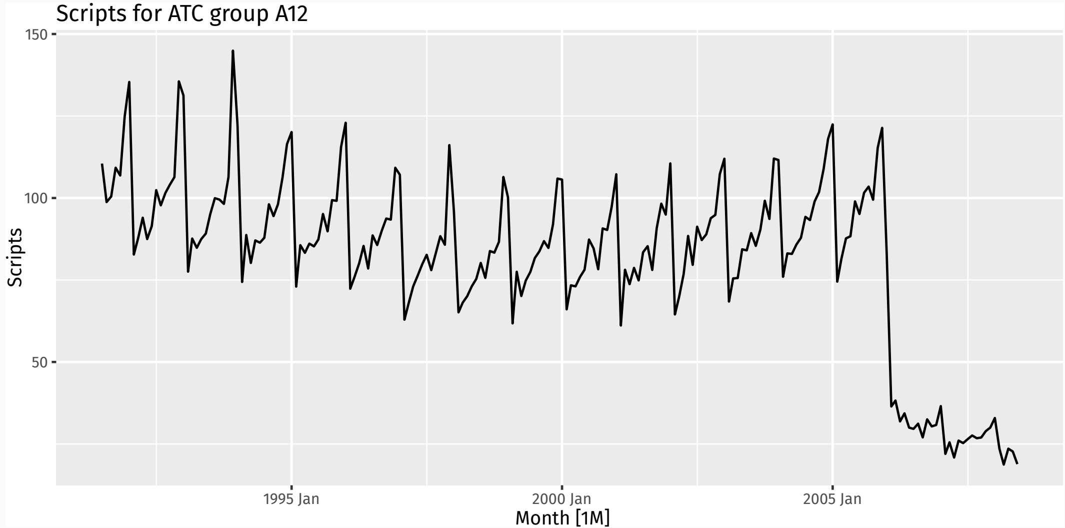
3 July 2024

Australian PBS data

```
pbs
```

```
# A tsibble: 17,016 x 3 [1M]
# Key:       ATC2 [84]
   ATC2     Month Scripts
   <chr>    <mth>   <dbl>
 1 A01   1991 Jul    22.6
 2 A01   1991 Aug    20.4
 3 A01   1991 Sep    21.4
 4 A01   1991 Oct    23.7
 5 A01   1991 Nov    23.5
 6 A01   1991 Dec    26.3
 7 A01   1992 Jan    22.0
 8 A01   1992 Feb    16.4
 9 A01   1992 Mar    17.2
10 A01   1992 Apr    18.8
# i 17,006 more rows
```

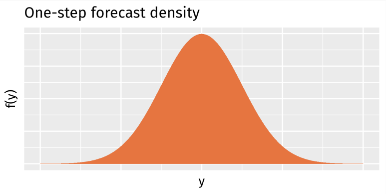
Australian PBS data



Anomaly score distribution

One-step forecast distribution: $N(\mu_t, \sigma^2)$

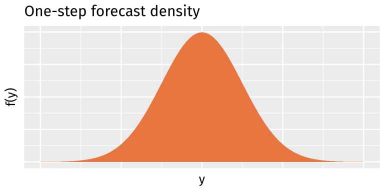
$$f(y_t|y_1,\dots,y_{t-1}) = \phi\left(\frac{y_t - \mu_t}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{\frac{(y_t-\mu_t)^2}{\sigma^2}\right\}$$



Anomaly score distribution

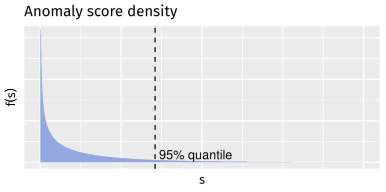
One-step forecast distribution: $N(\mu_t, \sigma^2)$

$$f(y_t|y_1,\dots,y_{t-1}) = \phi\left(\frac{y_t - \mu_t}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{\frac{(y_t-\mu_t)^2}{\sigma^2}\right\}$$



Anomaly score distribution: $S \sim \frac{1}{2}\chi^2_1 + c$

$$s_t = -\log f(y_t|y_1,\dots,y_{t-1}) = \frac{1}{2}\left(\frac{y_t-\mu_t}{2\sigma}\right)^2 + \frac{1}{2}\log(2\pi\sigma^2)$$



Anomaly score distribution

One-step forecast distribution: $N(\mu_t, \sigma^2)$

$$f(y_t|y_1,\dots,y_{t-1}) = \phi\left(\frac{y_t - \mu_t}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{\frac{(y_t-\mu_t)^2}{\sigma^2}\right\}$$

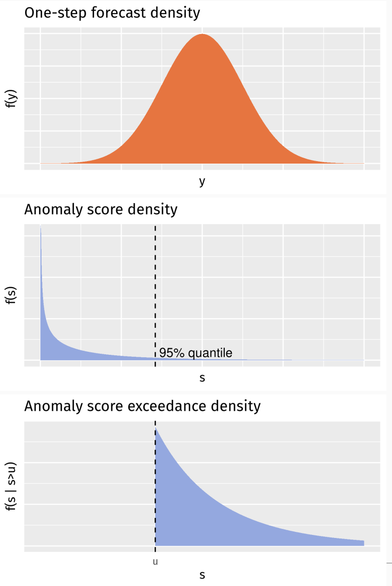
Anomaly score distribution: $S \sim \frac{1}{2}\chi_1^2 + c$

$$s_t = -\log f(y_t|y_1,\dots,y_{t-1}) = \frac{1}{2}\left(\frac{y_t-\mu_t}{2\sigma}\right)^2 + \frac{1}{2}\log(2\pi\sigma^2)$$

Extreme anomaly score distribution

$$H(x) = P(S \le u + x \mid S > u)$$

→ Generalized Pareto Distribution for almost all forecast distributions f.



Anomaly detection algorithm

For each t:

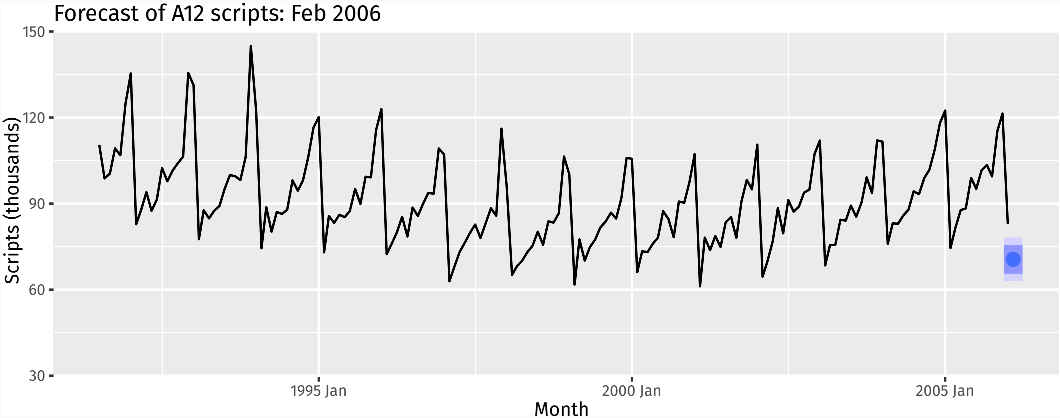
- Estimate one-step forecast density: $f(y_t|y_1,\dots,y_{t-1})$.
- Anomaly score: $s_t = -\log \hat{f}(y_t|y_1,\dots,y_{t-1})$.
- High anomaly score indicates potential anomaly.
- Fit a Generalized Pareto Distribution to the top 5% of anomaly scores seen so far.
- y_t is anomaly if $P(S > s_t) < 0.01$ under GPD.

Example

```r
a12 <- pbs |> filter(ATC2 == "A12", Month <= yearmonth("2006 Jan"))
a12plus <- pbs |> filter(ATC2 == "A12", Month <= yearmonth("2006 Feb"))
fc <- a12 |> model(ets = ETS(Scripts)) |> forecast(h = 1)
```

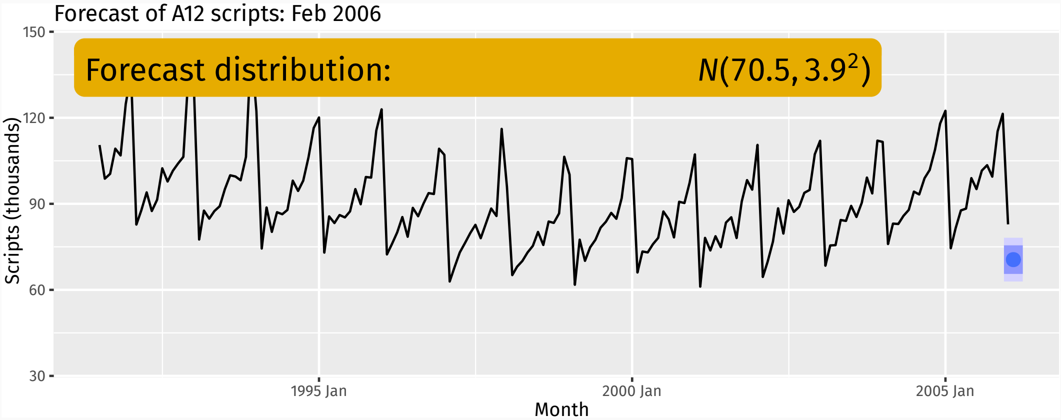
Example

```
a12 ← pbs ▷ filter(ATC2 == "A12", Month <= yearmonth("2006 Jan"))
a12plus ← pbs ▷ filter(ATC2 == "A12", Month <= yearmonth("2006 Feb"))
fc ← a12 ▷ model(ets = ETS(Scripts)) ▷ forecast(h = 1)
fc ▷ autoplot(a12)
```



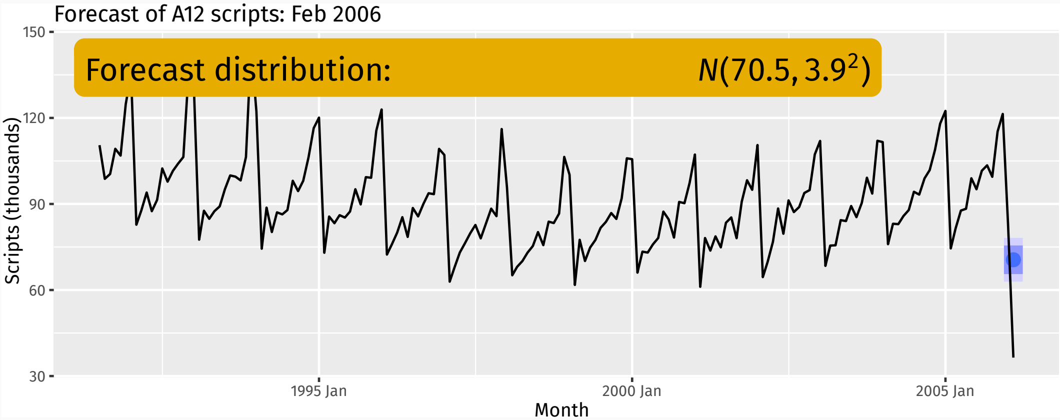
Example

```
a12 ← pbs ▷ filter(ATC2 == "A12", Month <= yearmonth("2006 Jan"))
a12plus ← pbs ▷ filter(ATC2 == "A12", Month <= yearmonth("2006 Feb"))
fc ← a12 ▷ model(ets = ETS(Scripts)) ▷ forecast(h = 1)
fc ▷ autoplot(a12)
```



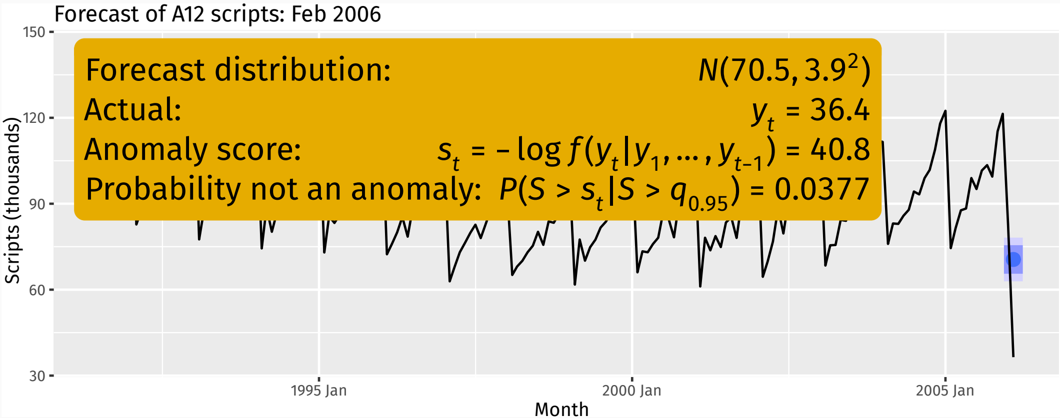
Example

```
a12 <- pbs |> filter(ATC2 == "A12", Month <= yearmonth("2006 Jan"))
a12plus <- pbs |> filter(ATC2 == "A12", Month <= yearmonth("2006 Feb"))
fc <- a12 |> model(ets = ETS(Scripts)) |> forecast(h = 1)
fc |> autoplot(a12plus)
```



Example

```r
a12 <- pbs |> filter(ATC2 == "A12", Month <= yearmonth("2006 Jan"))
a12plus <- pbs |> filter(ATC2 == "A12", Month <= yearmonth("2006 Feb"))
fc <- a12 |> model(ets = ETS(Scripts)) |> forecast(h = 1)
fc |> autoplot(a12plus)
```



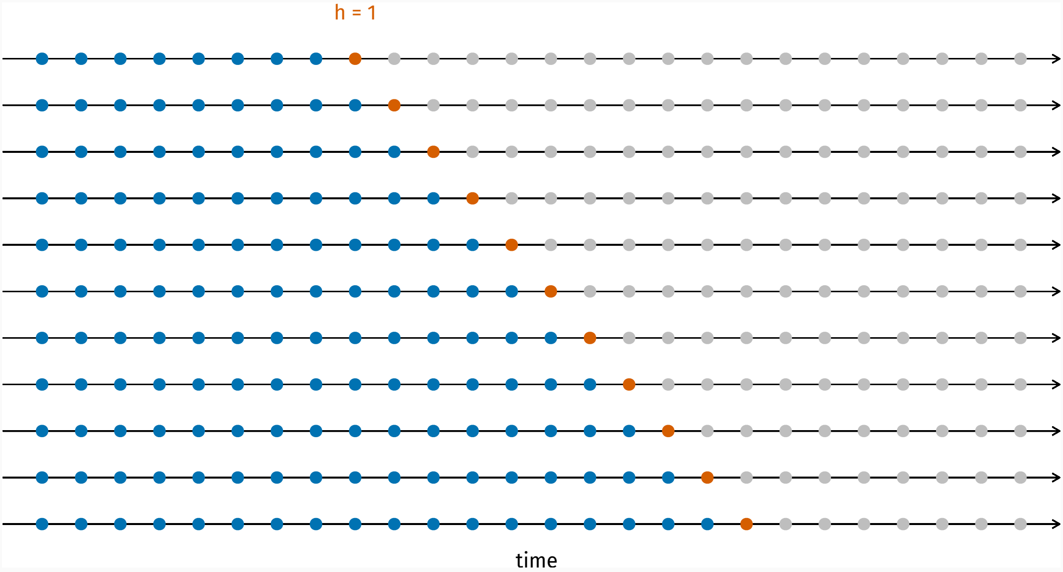
Forecast distribution: $N(70.5, 3.9^2)$

Actual: $y_t = 36.4$

Anomaly score: $s_t = -\log f(y_t|y_1, \dots, y_{t-1}) = 40.8$

Probability not an anomaly: $P(S > s_t|S > q_{0.95}) = 0.0377$

Rolling origin forecasts



Rolling origin forecasts

```r
pbs_stretch <- stretch_tsibble(pbs, .step = 1, .init = 36)
```

```
# A tsibble: 1,684,884 x 4 [1M]
# Key:       .id, ATC2 [14,076]
   ATC2      Month Scripts   .id
   <chr>     <mth>   <dbl> <int>
 1 A01    1991 Jul    22.6     1
 2 A01    1991 Aug    20.4     1
 3 A01    1991 Sep    21.4     1
 4 A01    1991 Oct    23.7     1
 5 A01    1991 Nov    23.5     1
 6 A01    1991 Dec    26.3     1
 7 A01    1992 Jan    22.0     1
 8 A01    1992 Feb    16.4     1
 9 A01    1992 Mar    17.2     1
10 A01    1992 Apr    18.8     1
# i 1,684,874 more rows
```

Rolling origin forecasts

```
pbs_fit <- pbs_stretch |> model(ets = ETS(Scripts))
```

```
# A mable: 14,076 x 3
# Key:     .id, ATC2 [14,076]
     .id ATC2             ets
   <int> <chr>        <model>
 1     1 A01    <ETS(M,N,A)>
 2     1 A02    <ETS(M,A,M)>
 3     1 A03    <ETS(M,A,M)>
 4     1 A04    <ETS(M,N,A)>
 5     1 A05   <ETS(A,Ad,N)>
 6     1 A06    <ETS(M,A,M)>
 7     1 A07    <ETS(M,N,M)>
 8     1 A09    <ETS(M,A,M)>
 9     1 A10    <ETS(M,A,M)>
10     1 A11    <ETS(M,A,M)>
# i 14,066 more rows
```

Rolling origin forecasts

```r
pbs_fc <- forecast(pbs_fit, h = 1)
```

```
# A fable: 14,076 x 4 [1M]
# Key:     .id, ATC2 [14,076]
     .id ATC2     Month        Scripts
   <int> <chr>    <mth>         <dist>
 1     1 A01   1994 Jul    N(23, 2.1)
 2     1 A02   1994 Jul  N(590, 1054)
 3     1 A03   1994 Jul      N(84, 19)
 4     1 A04   1994 Jul      N(69, 15)
 5     1 A05   2003 Jul N(1.4, 0.014)
 6     1 A06   1994 Jul    N(33, 4.2)
 7     1 A07   1994 Jul      N(74, 17)
 8     1 A09   1994 Jul N(3.7, 0.029)
 9     1 A10   1994 Jul    N(166, 54)
10     1 A11   1994 Jul       N(30, 3)
# i 14,066 more rows
```

PBS anomalies

```r
pbs_scores <- pbs_fc |>
  left_join(pbs |> rename(actual = Scripts), by = c("ATC2", "Month")) |>
  mutate(
    s = -log_likelihood(Scripts, actual), # Density scores
    prob = lookout(density_scores = s)     # Probability not an anomaly
  )
```

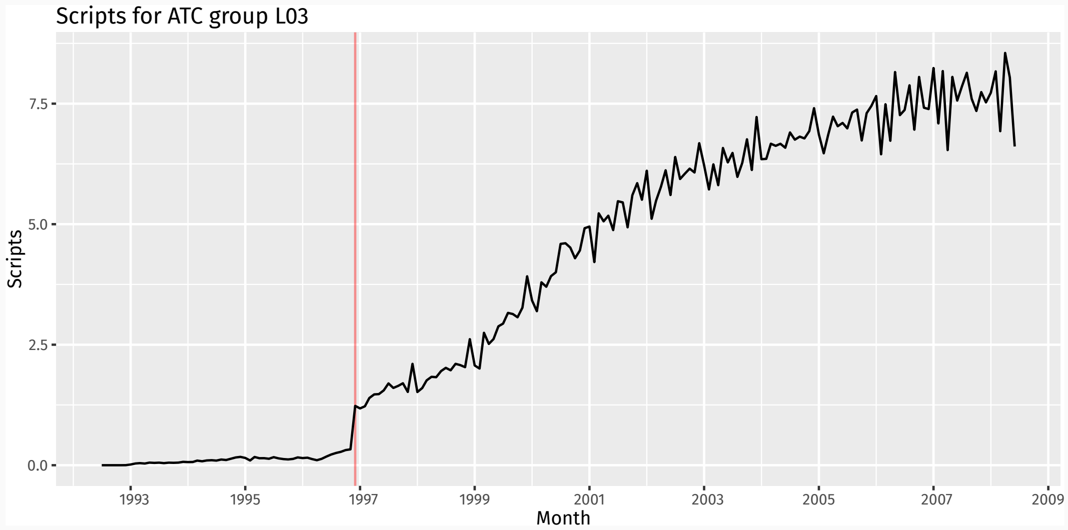
```
# A fable: 14,076 x 7 [1M]
# Key:     .id, ATC2 [14,076]
     .id ATC2     Month        Scripts actual     s  prob
   <int> <chr>    <mth>         <dist>  <dbl> <dbl> <dbl>
 1     1 A01   1994 Jul    N(23, 2.1)  20.9   2.46 1
 2     1 A02   1994 Jul  N(590, 1054) 516.    6.97 0.554
 3     1 A03   1994 Jul     N(84, 19)  80.5   2.75 1
 4     1 A04   1994 Jul     N(69, 15)  66.1   2.62 1
 5     1 A05   2003 Jul N(1.4, 0.014)   1.47 -1.05 1
 6     1 A06   1994 Jul    N(33, 4.2)  29.2   3.41 1
 7     1 A07   1994 Jul     N(74, 17)  68.5   3.09 1
 8     1 A09   1994 Jul N(3.7, 0.029)   3.32  1.46 1
```

PBS anomalies

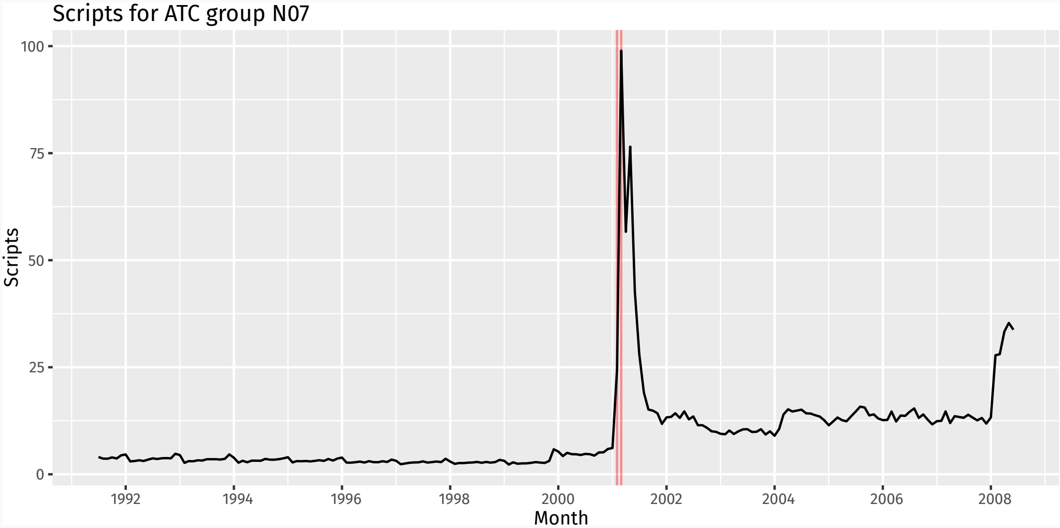
```r
pbs_scores |> filter(prob < 0.01)
```

```
# A fable: 12 x 7 [1M]
# Key:     .id, ATC2 [12]
     .id ATC2     Month                Scripts actual     s     prob
   <int> <chr>    <mth>                 <dist>  <dbl> <dbl>    <dbl>
 1    18 L03   1996 Dec    N(0.33, 0.00054)  1.23   756. 0.00194
 2    21 C05   1996 Mar N(-0.0099, 5.2e-06)  0.04   236. 0.00613
 3    24 C05   1996 Jun   N(0.005, 1.8e-06)  0.05   560. 0.00260
 4    33 G01   1997 Mar          N(47, 6.3)  3.60   154. 0.00942
 5    42 D     1997 Dec       N(4.4, 0.055)  0.407  145. 0.00995
 6    44 R06   1998 Feb     N(-0.97, 0.011)  4.99  1623. 0.000916
 7    55 R06   1999 Jan      N(-1.1, 0.019)  1.45   168. 0.00859
 8    80 N07   2001 Feb       N(4.3, 0.14) 24.6   1469. 0.00101
 9    81 N07   2001 Mar          N(10, 6.8) 98.9    582. 0.00251
10   131 D11   2005 May    N(0.13, 0.00017)  0.596  608. 0.00240
11   141 P01   2006 Mar    N(0.18, 0.00022)  1.50  3882. 0.000390
12   146 P01   2006 Aug   N(0.013, 1.5e-06)  0.129 4607. 0.000330
```

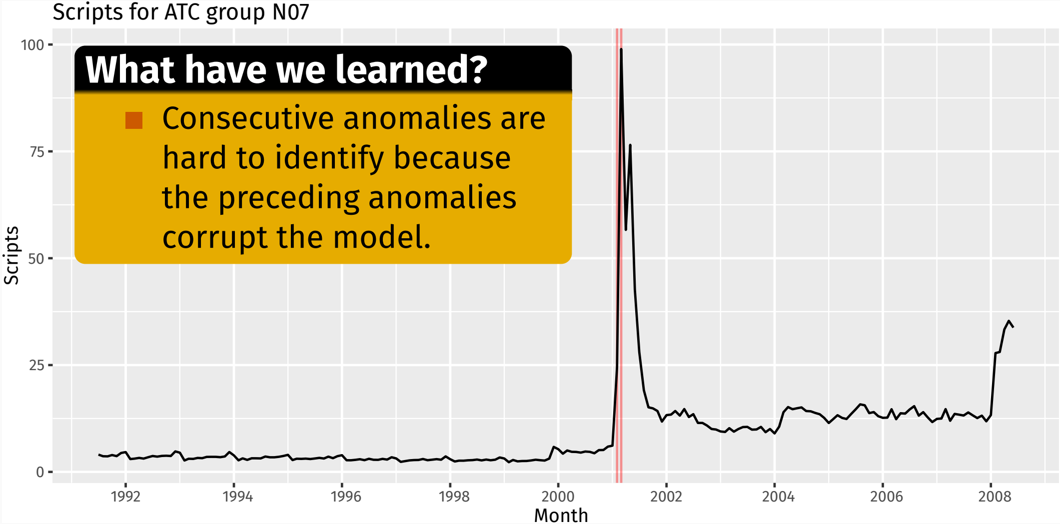
PBS anomalies



PBS anomalies



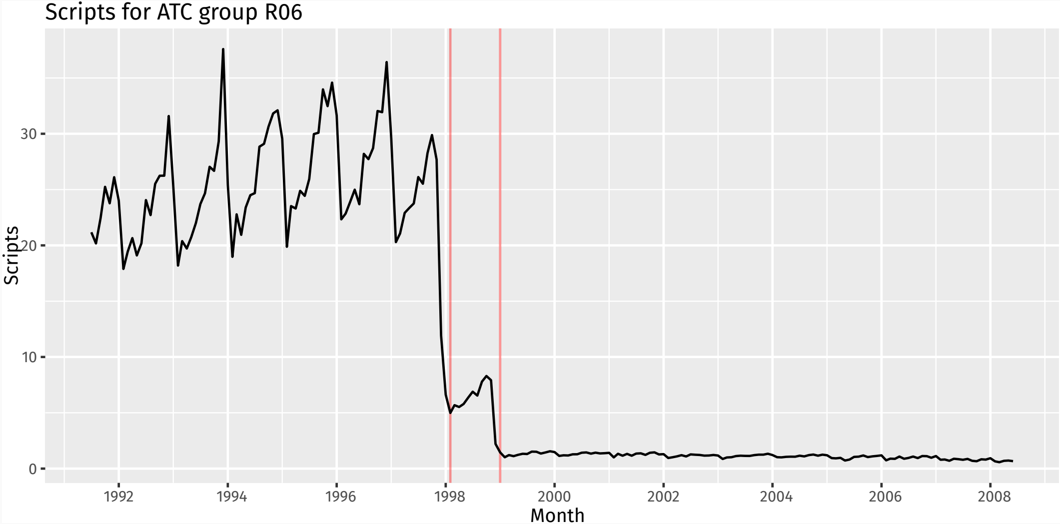
PBS anomalies



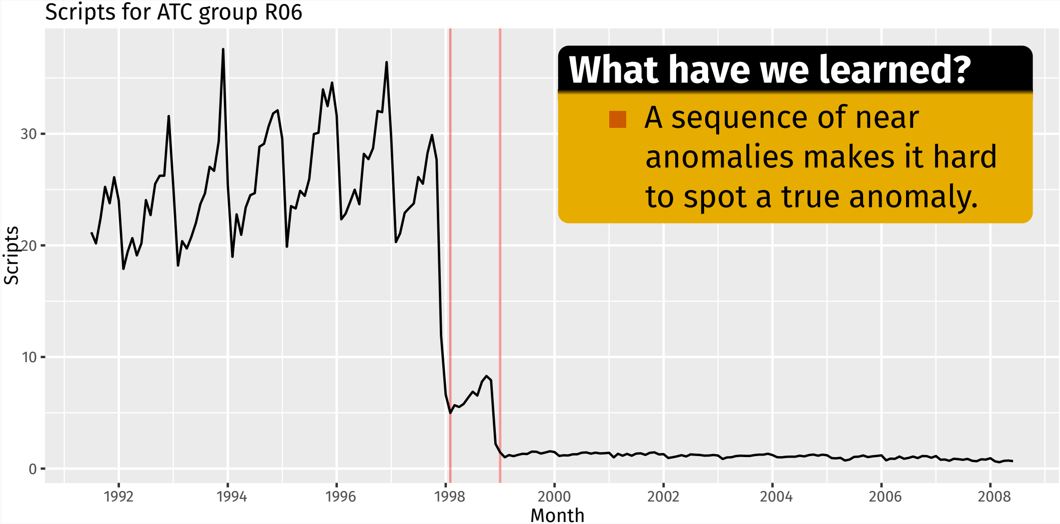
What have we learned?

- Consecutive anomalies are hard to identify because the preceding anomalies corrupt the model.

PBS anomalies



PBS anomalies



What have we learned?

- A sequence of near anomalies makes it hard to spot a true anomaly.

Modified anomaly detection algorithm

For each t:

- Estimate one-step forecast density: $f(y_t|y_1,\dots,y_{t-1})$.
- Anomaly score: $s_t = -\log \hat{f}(y_t|y_1,\dots,y_{t-1})$.
- High anomaly score indicates potential anomaly.
- Fit a Generalized Pareto Distribution to the top 5% of anomaly scores seen so far.
- y_t is anomaly if $P(S > s_t) < 0.01$ under GPD.
- **If y_t is anomaly, set y_t to missing for next iteration.**

Example: French mortality

```
fr_mortality
```

```
# A tsibble: 41,612 x 4 [1Y]
# Key:       Age, Sex [202]
    Year   Age Sex    Mortality
   <int> <int> <chr>      <dbl>
 1  1816     0 Female     0.187
 2  1817     0 Female     0.182
 3  1818     0 Female     0.186
 4  1819     0 Female     0.197
 5  1820     0 Female     0.181
 6  1821     0 Female     0.182
 7  1822     0 Female     0.207
 8  1823     0 Female     0.192
 9  1824     0 Female     0.199
10  1825     0 Female     0.194
# i 41,602 more rows
```

Example: French mortality

```r
fr_stretch <- fr_mortality |> stretch_tsibble(.init = 30, .step=1)
```

```
# A tsibble: 4,218,972 x 5 [1Y]
# Key:       .id, Age, Sex [35,754]
     .id  Year   Age Sex    Mortality
   <int> <int> <int> <chr>      <dbl>
 1     1  1816     0 Female     0.187
 2     1  1817     0 Female     0.182
 3     1  1818     0 Female     0.186
 4     1  1819     0 Female     0.197
 5     1  1820     0 Female     0.181
 6     1  1821     0 Female     0.182
 7     1  1822     0 Female     0.207
 8     1  1823     0 Female     0.192
 9     1  1824     0 Female     0.199
10     1  1825     0 Female     0.194
# i 4,218,962 more rows
```

Example: French mortality

```r
fit ← fr_stretch ▷ model(arima = ARIMA(Mortality))
fc ← forecast(fit, h = 1)
fr_scores ← fc ▷
  select(Year, Age, Sex, Mortality) ▷
  left_join(fr_mortality ▷ rename(actual = Mortality)) ▷
  mutate(
    s = -log_likelihood(Mortality, actual), # Density scores
    prob = lookout(density_scores = s)    # Probability not an anomaly
  )
```

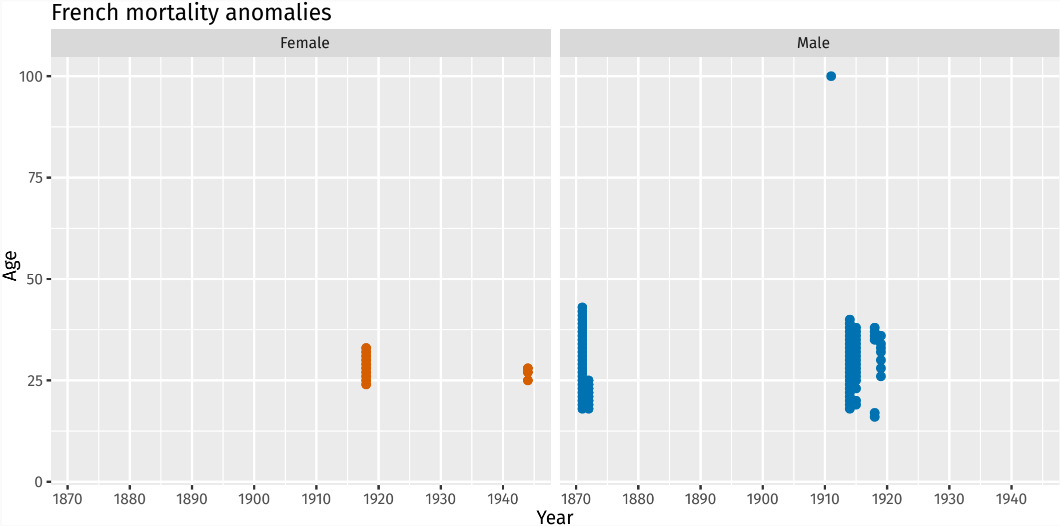
```
# A fable: 35,754 x 9 [1Y]
# Key:     Age, Sex, .id, .model [35,754]
    Year   Age Sex              Mortality   .id .model actual     s  prob
   <dbl> <int> <chr>               <dist> <int> <chr>   <dbl> <dbl> <dbl>
 1  1846     0 Female  N(0.15, 0.00016)     1 arima  0.167  -2.70     1
 2  1846     0 Male    N(0.18, 0.00021)     1 arima  0.195  -2.48     1
 3  1846     1 Female N(0.057, 3.6e-05)     1 arima  0.0550 -4.11     1
 4  1846     1 Male   N(0.058, 3.6e-05)     1 arima  0.0555 -4.09     1
 5  1846     2 Female  N(0.04, 1.5e-05)     1 arima  0.0398 -4.61     1
```

Example: French mortality

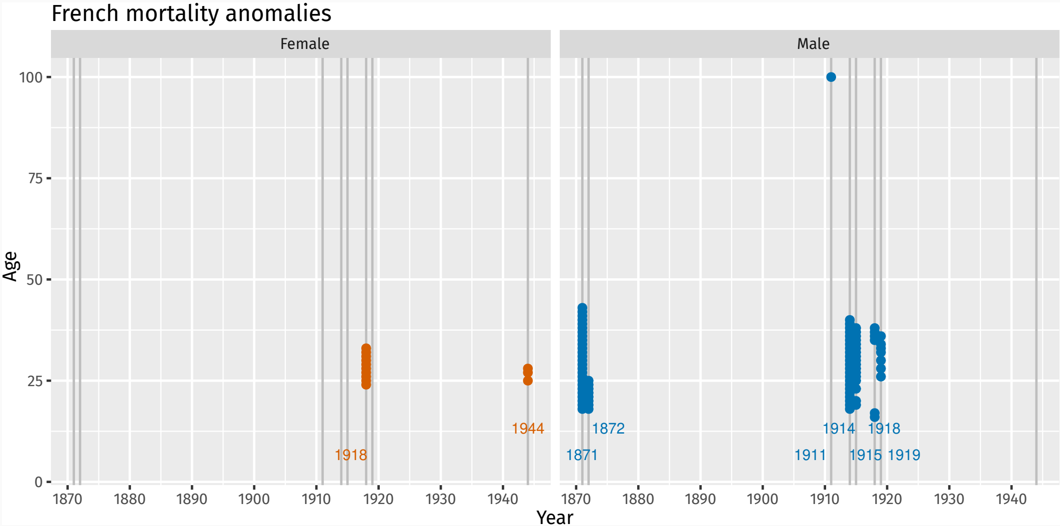
```
fr_scores |> arrange(prob)
```

```
# A tsibble: 35,754 x 9 [1Y]
# Key:       Age, Sex, .id, .model [35,754]
    Year   Age Sex             Mortality   .id .model actual     s     prob
   <dbl> <int> <chr>              <dist> <int> <chr>   <dbl> <dbl>    <dbl>
 1  1914    18 Male  N(0.0055, 1.4e-06)    69 arima  0.0798 1965. 0.000902
 2  1914    19 Male  N(0.0063, 4.1e-06)    69 arima  0.0906  872. 0.00194
 3  1914    29 Male  N(0.0075, 2.5e-06)    69 arima  0.0597  549. 0.00301
 4  1914    30 Male  N(0.0083, 2.4e-06)    69 arima  0.0591  544. 0.00304
 5  1914    31 Male  N(0.0086, 2.4e-06)    69 arima  0.0578  489. 0.00336
 6  1914    28 Male  N(0.0074, 2.9e-06)    69 arima  0.0611  485. 0.00338
 7  1914    32 Male  N(0.0087, 2.4e-06)    69 arima  0.0550  439. 0.00371
 8  1914    27 Male    N(0.0073, 4e-06)    69 arima  0.0613  356. 0.00452
 9  1914    33 Male   N(0.009, 2.3e-06)    69 arima  0.0493  345. 0.00466
10  1914    26 Male  N(0.0073, 6.1e-06)    69 arima  0.0648  266. 0.00595
# i 35,744 more rows
```

Example: French mortality



Example: French mortality



Example: French mortality