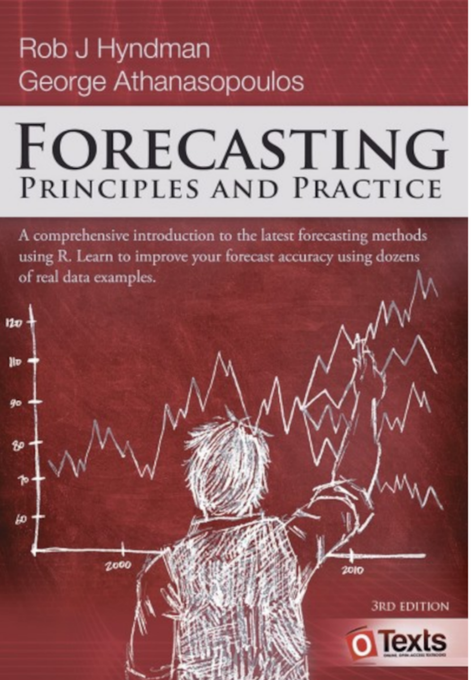


9. ARIMA models

9.1 Stationarity and differencing

OTexts.org/fpp3/

Stationarity

Definition

If $\{y_t\}$ is a stationary time series, then for all s, the distribution of $(y_t, \ldots, y_{t+s})$ does not depend on t.

Stationarity

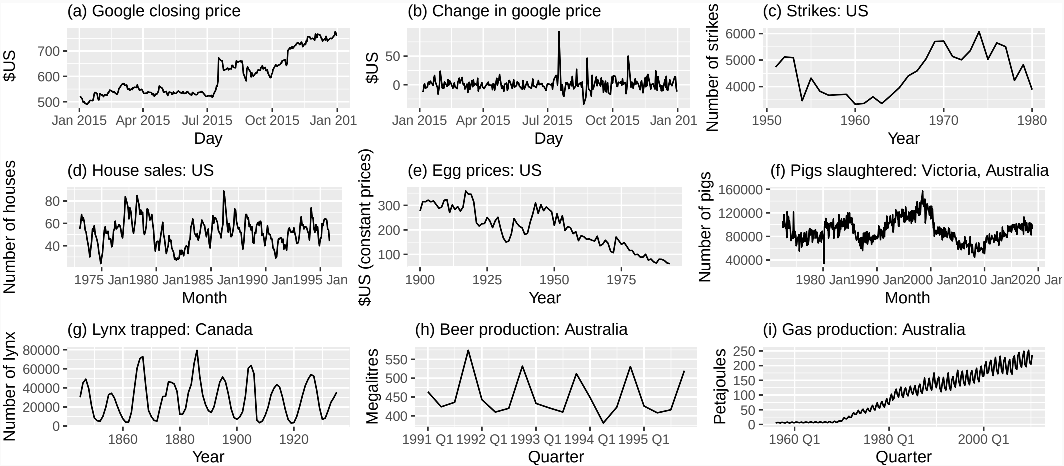
Definition

If $\{y_t\}$ is a stationary time series, then for all s, the distribution of $(y_t, \ldots, y_{t+s})$ does not depend on t.

A stationary series is:

- roughly horizontal
- constant variance
- no patterns predictable in the long-term

Stationary or not



Stationarity

Definition

If $\{y_t\}$ is a stationary time series, then for all s, the distribution of $(y_t, \ldots, y_{t+s})$ does not depend on t.

Stationarity

Definition

If $\{y_t\}$ is a stationary time series, then for all s, the distribution of $(y_t, \dots, y_{t+s})$ does not depend on t.

- Transformations help to stabilize the variance.
- For ARIMA modelling, we also need to stabilize the mean.

Non-stationarity in the mean

Identifying non-stationary series

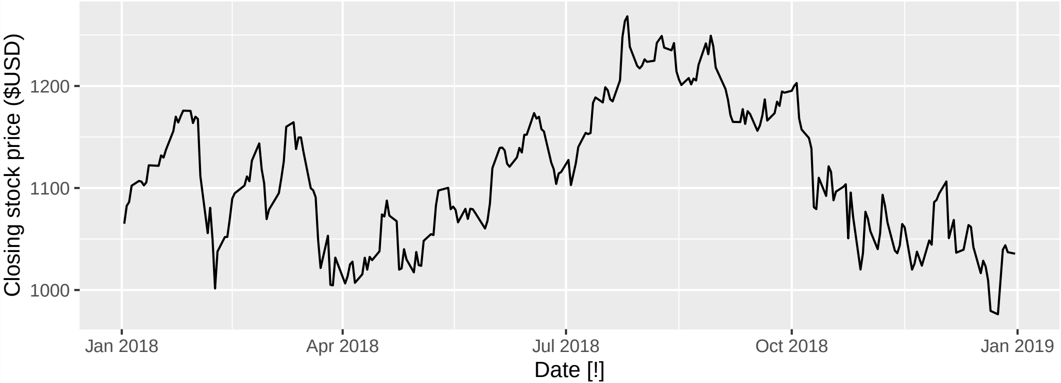
- time plot.
- The ACF of stationary data drops to zero relatively quickly
- The ACF of non-stationary data decreases slowly.
- For non-stationary data, the value of r_1 is often large and positive.

Example: Google stock price

```r
google_2018 <- gafa_stock |>
  filter(Symbol == "GOOG", year(Date) == 2018)
```

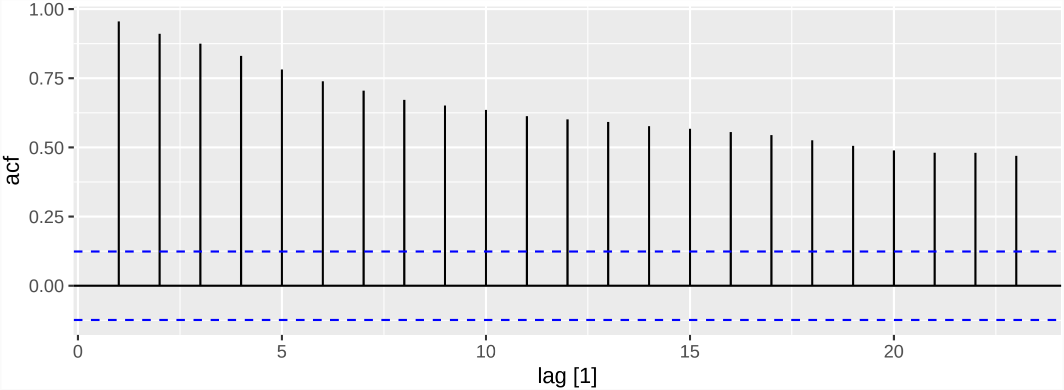
Example: Google stock price

```r
google_2018 |>
  autoplot(Close) +
  labs(y = "Closing stock price ($USD)")
```



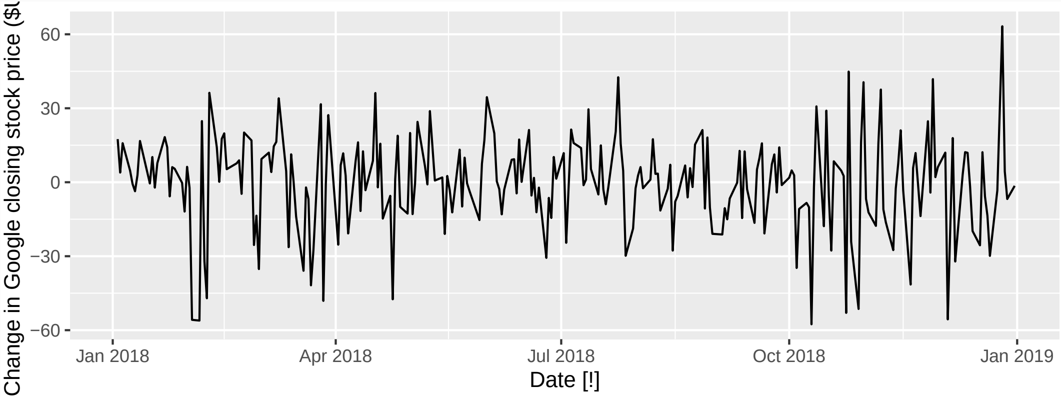
Example: Google stock price

```
google_2018 |>
  ACF(Close) |>
  autoplot()
```



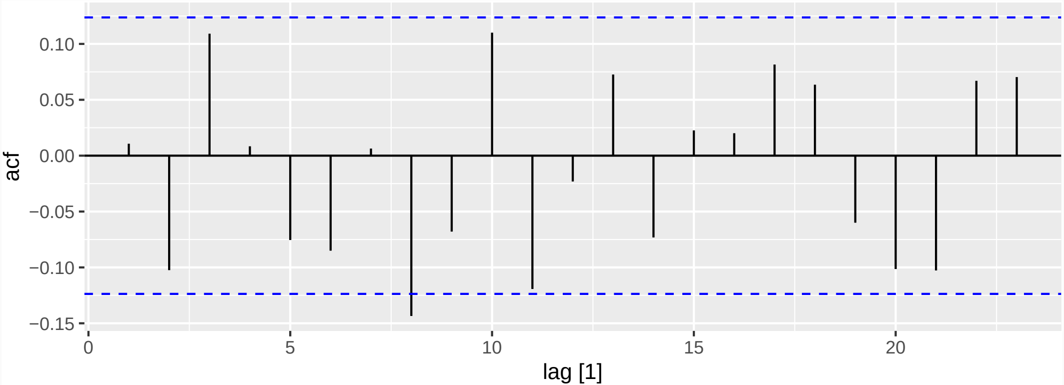
Example: Google stock price

```r
google_2018 |>
  autoplot(difference(Close)) +
  labs(y = "Change in Google closing stock price ($USD)")
```



Example: Google stock price

```r
google_2018 |>
  ACF(difference(Close)) |>
  autoplot()
```



Differencing

- Differencing helps to stabilize the mean.
- The differenced series is the change between each observation in the original series: $y'_t = y_t - y_{t-1}$.
- The differenced series will have only $T - 1$ values since it is not possible to calculate a difference y'_1 for the first observation.

Second-order differencing

Occasionally the differenced data will not appear stationary and it may be necessary to difference the data a second time:

Second-order differencing

Occasionally the differenced data will not appear stationary and it may be necessary to difference the data a second time:

$$\begin{aligned} y''_t &= y'_t - y'_{t-1} \\ &= (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) \\ &= y_t - 2y_{t-1} + y_{t-2}. \end{aligned}$$

Second-order differencing

Occasionally the differenced data will not appear stationary and it may be necessary to difference the data a second time:

$$\begin{aligned} y''_t &= y'_t - y'_{t-1} \\ &= (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) \\ &= y_t - 2y_{t-1} + y_{t-2}. \end{aligned}$$

- y''_t will have $T - 2$ values.
- In practice, it is almost never necessary to go beyond second-order differences.

Seasonal differencing

A seasonal difference is the difference between an observation and the corresponding observation from the previous year.

$$y'_t = y_t - y_{t-m}$$

where m = number of seasons.

Seasonal differencing

A seasonal difference is the difference between an observation and the corresponding observation from the previous year.

$$y'_t = y_t - y_{t-m}$$

where m = number of seasons.

- For monthly data $m = 12$.
- For quarterly data $m = 4$.
- Seasonally differenced series will have $T - m$ obs.

Seasonal differencing

A seasonal difference is the difference between an observation and the corresponding observation from the previous year.

$$y'_t = y_t - y_{t-m}$$

where m = number of seasons.

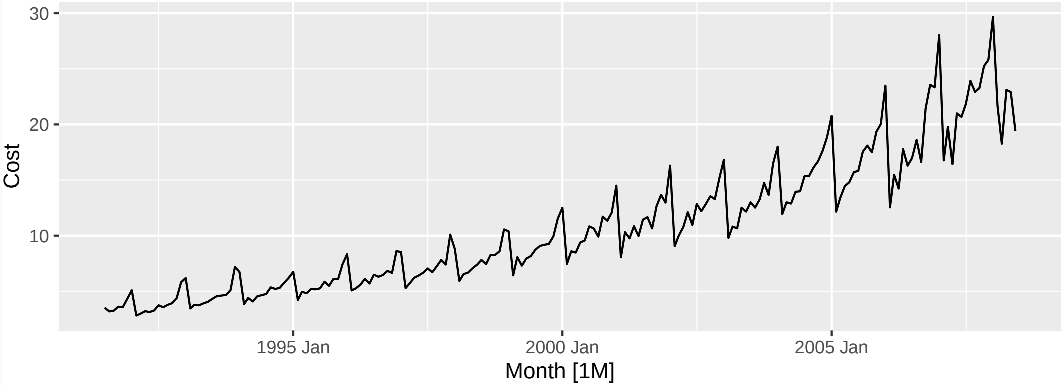
- For monthly data $m = 12$.
- For quarterly data $m = 4$.
- Seasonally differenced series will have $T - m$ obs.

Antidiabetic drug sales

```r
a10 <- PBS |>
  filter(ATC2 == "A10") |>
  summarise(Cost = sum(Cost) / 1e6)
```

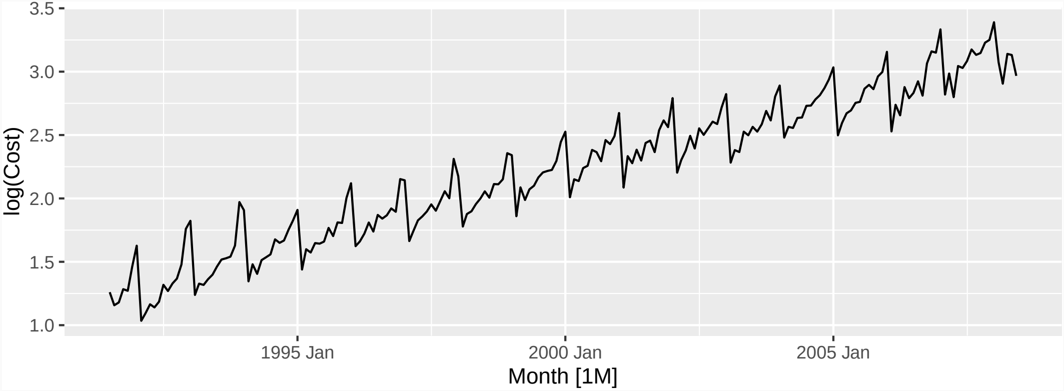
Antidiabetic drug sales

```r
a10 |> autoplot(
  Cost
)
```



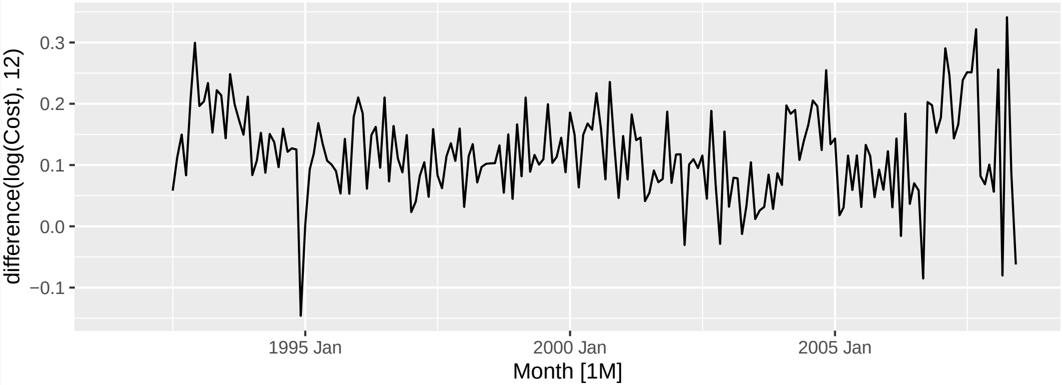
Antidiabetic drug sales

```r
a10 |> autoplot(
  log(Cost)
)
```



Antidiabetic drug sales

```r
a10 |> autoplot(
  log(Cost) |> difference(12)
)
```

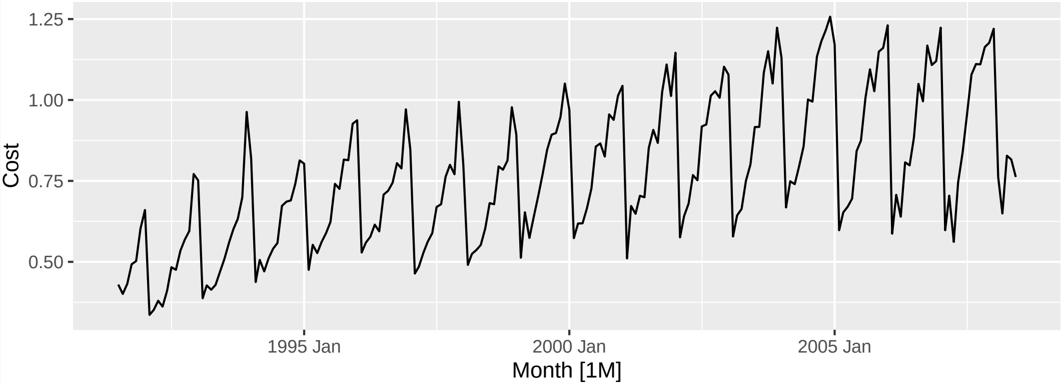


Corticosteroid drug sales

```r
h02 <- PBS |>
  filter(ATC2 == "H02") |>
  summarise(Cost = sum(Cost) / 1e6)
```

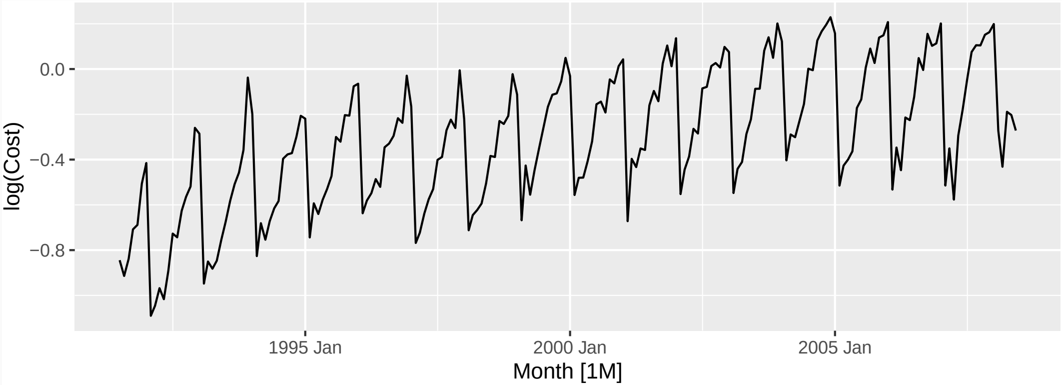
Corticosteroid drug sales

```r
h02 |> autoplot(
  Cost
)
```



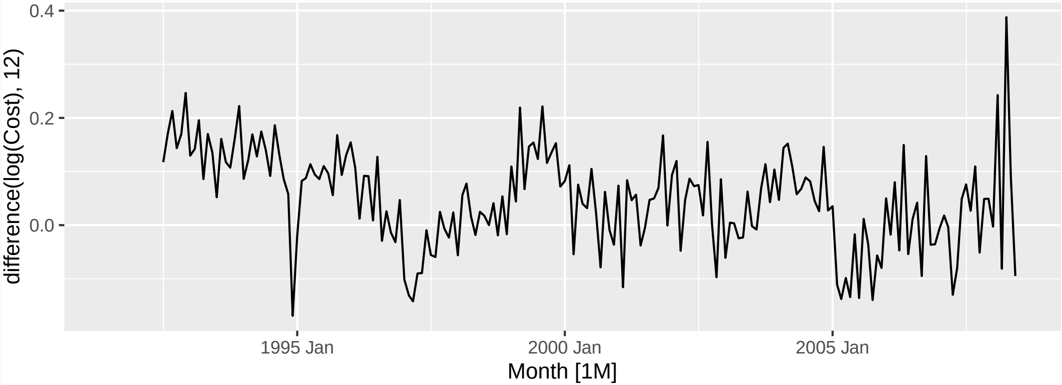
Corticosteroid drug sales

```r
h02 |> autoplot(
  log(Cost)
)
```



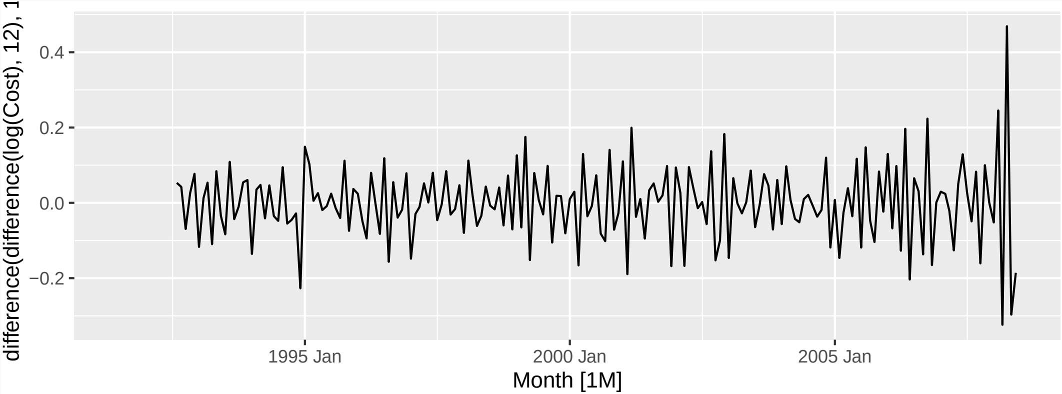
Corticosteroid drug sales

```r
h02 |> autoplot(
  log(Cost) |> difference(12)
)
```



Corticosteroid drug sales

```r
h02 |> autoplot(
  log(Cost) |> difference(12) |> difference(1)
)
```



Corticosteroid drug sales

- Seasonally differenced series is closer to being stationary.
- Remaining non-stationarity can be removed with further first difference.

If $y'_t = y_t - y_{t-12}$ denotes seasonally differenced series, then twice-differenced series is

$$\begin{aligned} y^*_t &= y'_t - y'_{t-1} \\ &= (y_t - y_{t-12}) - (y_{t-1} - y_{t-13}) \\ &= y_t - y_{t-1} - y_{t-12} + y_{t-13}\,. \end{aligned}$$

Seasonal differencing

When both seasonal and first differences are applied…

Seasonal differencing

When both seasonal and first differences are applied…

- it makes no difference which is done first—the result will be the same.
- If seasonality is strong, we recommend that seasonal differencing be done first because sometimes the resulting series will be stationary and there will be no need for further first difference.

Seasonal differencing

When both seasonal and first differences are applied...

- it makes no difference which is done first—the result will be the same.
- If seasonality is strong, we recommend that seasonal differencing be done first because sometimes the resulting series will be stationary and there will be no need for further first difference.

It is important that if differencing is used, the differences are interpretable.

Interpretation of differencing

- first differences are the change between one observation and the next;
- seasonal differences are the change between one year to the next.

Interpretation of differencing

- first differences are the change between one observation and the next;
- seasonal differences are the change between one year to the next.

But taking lag 3 differences for yearly data, for example, results in a model which cannot be sensibly interpreted.