2.3 Stationarity and differencing
1 Stationarity
2 Differencing
3 Unit root tests
4 Lab session 15
5 Backshift notation
Stationarity

**Definition**

<table>
<thead>
<tr>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>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 ).</td>
</tr>
</tbody>
</table>
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?

![Dow Jones Index Chart]
Stationary?

Change in Dow Jones Index

Day

-100
-50
0
50
0
50
100
150
200
250
300

-100
-50
0
50
0
50
100
150
200
250
300

0
50
100
150
200
250
300
400
500

Stationary?
Stationary?

Price of a dozen eggs in 1993 dollars

Year

$
Number of pigs slaughtered in Victoria

Stationary?
Stationary?

Annual Canadian Lynx Trappings

Year
Number trapped
1820 1840 1860 1880 1900 1920
0 2000 4000 6000
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 \( t \).

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: Dow-Jones index

```r
autoplot(dj) + ylab("Dow Jones Index") + xlab("Day")
```
Example: Dow-Jones index

$\text{ggAcf}(\text{dj})$
Example: Dow-Jones index

```r
autoplot(diff(dj)) +
  ylab("Change in Dow Jones Index") + xlab("Day")
```
Example: Dow-Jones index

`ggAcf(diff(dj))`
Outline

1. Stationarity
2. Differencing
3. Unit root tests
4. Lab session 15
5. Backshift notation
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.
Occasionally the differenced data will not appear stationary and it may be necessary to difference the data a second time:

\[ y''_t = y'_t - y'_{t-1} = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2}. \]
Occasionally the differenced data will not appear stationary and it may be necessary to difference the data a second time:

\[ y''_t = y'_{t} - y'_{t-1} \]

\[ = (y_{t} - y_{t-1}) - (y_{t-1} - y_{t-2}) \]

\[ = y_{t} - 2y_{t-1} + y_{t-2}. \]
Occasionally the differenced data will not appear stationary and it may be necessary to difference the data a second time:

\[ y''_t = y'_t - y'_{t-1} \]

\[ = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) \]

\[ = y_t - 2y_{t-1} + y_{t-2}. \]

- \( y''_t \) will have \( T - 2 \) values.
- In practice, it is almost never necessary to go beyond second-order differences.
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 \( m \) is the number of seasons.

For monthly data \( m = 12 \).

For quarterly data \( m = 4 \).
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 \( m = 4 \).
Electricity production

```r
usmelec %>% autoplot()
```
Electricity production

\texttt{usmelec} \( \%\%\% \texttt{log(\() \%\%\% \texttt{autoplot(\()}

\begin{center}
\begin{tikzpicture}
\begin{axis}[
    width=\textwidth,
    height=\textwidth,
    xmin=1980, xmax=2010,
    ymin=5.1, ymax=6.0,
    ytick={5.1,5.4,5.7,6.0},
    yticklabels={5.1,5.4,5.7,6.0},
    xlabel={Time},
    ylabel={},
    grid=major
]\end{axis}\end{tikzpicture}
\end{center}
Electricity production

usmelec %>% log() %>% diff(lag=12) %>% autoplot()
Electricity production

\texttt{usmelec >>\% log() >>\% diff(lag=12) >>\% diff(lag=1) >>\% autoplot()}

![Graph showing electricity production over time.](image)
Electricity production

- 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

\[
y^*_t = y'_t - y'_{t-1} \\
= (y_t - y_{t-12}) - (y_{t-1} - y_{t-13}) \\
= y_t - y_{t-1} - y_{t-12} + y_{t-13}
\]
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.
Outline

1. Stationarity
2. Differencing
3. Unit root tests
4. Lab session 15
5. Backshift notation
Unit root tests

Statistical tests to determine the required order of differencing.

1. Augmented Dickey Fuller test: null hypothesis is that the data are non-stationary and non-seasonal.
2. Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test: null hypothesis is that the data are stationary and non-seasonal.
3. Other tests available for seasonal data.
library(urca)
summary(ur.kpss(goog))

## # KPSS Unit Root Test #
## # KPSS Unit Root Test #
## Test is of type: mu with 7 lags.
## Value of test-statistic is: 10.72
## Critical value for a significance level of:
## 10pct  5pct  2.5pct  1pct
## critical values 0.347  0.463  0.574  0.739
KPSS test

```r
library(urca)
summary(ur.kpss(goog))
```

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 7 lags.
##
## Value of test-statistic is: 10.72
##
## Critical value for a significance level of:
## 10pct  5pct  2.5pct  1pct
## critical values  0.347  0.463  0.574  0.739
```

```r
ndiffs(goog)
```

```
## [1] 1
```
Automatically selecting differences

STL decomposition: \( y_t = T_t + S_t + R_t \)

Seasonal strength \( F_s = \max \left( 0, 1 - \frac{\text{Var}(R_t)}{\text{Var}(S_t+R_t)} \right) \)

If \( F_s > 0.64 \), do one seasonal difference.
Automatically selecting differences

STL decomposition: \( y_t = T_t + S_t + R_t \)

Seasonal strength \( F_s = \max \left( 0, 1 - \frac{\text{Var}(R_t)}{\text{Var}(S_t + R_t)} \right) \)

If \( F_s > 0.64 \), do one seasonal difference.

```
usmelec %>% log() %>% nsdiffs()
```

```
## [1] 1
```

```
usmelec %>% log() %>% diff(lag=12) %>% ndiffs()
```

```
## [1] 1
```
Outline

1. Stationarity
2. Differencing
3. Unit root tests
4. Lab session 15
5. Backshift notation
Lab Session 15
Outline

1. Stationarity
2. Differencing
3. Unit root tests
4. Lab session 15
5. Backshift notation
Backshift notation

A very useful notational device is the backward shift operator, $B$, which is used as follows:

$$By_t = y_{t-1}.$$
A very useful notational device is the backward shift operator, $B$, which is used as follows:

$$By_t = y_{t-1}.$$ 

In other words, $B$, operating on $y_t$, has the effect of shifting the data back one period.
A very useful notational device is the backward shift operator, $B$, which is used as follows:

$$B y_t = y_{t-1}.$$ 

In other words, $B$, operating on $y_t$, has the effect of shifting the data back one period. Two applications of $B$ to $y_t$ shifts the data back two periods:

$$B(By_t) = B^2 y_t = y_{t-2}.$$
A very useful notational device is the backward shift operator, $B$, which is used as follows:

$$By_t = y_{t-1}.$$ 

In other words, $B$, operating on $y_t$, has the effect of shifting the data back one period. Two applications of $B$ to $y_t$ shifts the data back two periods:

$$B(By_t) = B^2 y_t = y_{t-2}.$$ 

For monthly data, if we wish to shift attention to “the same month last year,” then $B^{12}$ is used, and the notation is $B^{12} y_t = y_{t-12}$. 

The backward shift operator is convenient for describing the process of differencing.
The backward shift operator is convenient for describing the process of *differencing*. A first difference can be written as

$$y'_t = y_t - y_{t-1} = y_t - By_t = (1 - B)y_t.$$
The backward shift operator is convenient for describing the process of differencing. A first difference can be written as

\[ y'_t = y_t - y_{t-1} = y_t - By_t = (1 - B)y_t. \]

Note that a first difference is represented by \((1 - B)\).
The backward shift operator is convenient for describing the process of differencing. A first difference can be written as

\[ y'_t = y_t - y_{t-1} = y_t - By_t = (1 - B)y_t. \]

Note that a first difference is represented by \((1 - B)\).

Similarly, if second-order differences (i.e., first differences of first differences) have to be computed, then:

\[ y''_t = y_t - 2y_{t-1} + y_{t-2} = (1 - B)^2y_t. \]
Second-order difference is denoted \((1 - B)^2\).

Second-order difference is not the same as a second difference, which would be denoted \(1 - B^2\);

In general, a \(d\)th-order difference can be written as

\[(1 - B)^d y_t.\]

A seasonal difference followed by a first difference can be written as

\[(1 - B)(1 - B^m)y_t.\]
The “backshift” notation is convenient because the terms can be multiplied together to see the combined effect.

$$(1 - B)(1 - B^m)y_t = (1 - B - B^m + B^{m+1})y_t$$

$$= y_t - y_{t-1} - y_{t-m} + y_{t-m-1}.$$
The “backshift” notation is convenient because the terms can be multiplied together to see the combined effect.

\[(1 - B)(1 - B^m)y_t = (1 - B - B^m + B^{m+1})y_t\]

\[= y_t - y_{t-1} - y_{t-m} + y_{t-m-1}.\]

For monthly data, \(m = 12\) and we obtain the same result as earlier.